

Finnish Cancer Registry
and
Department of Public Health, Faculty of Medicine, University of Helsinki

JOHANNA SEPPÄNEN

**Mammographic mass-screening and future breast
cancer burden in Finland**

ACADEMIC DISSERTATION

*To be presented, with the permission of the Faculty of Medicine, University of Helsinki, for
public examination in Meilahti Campus Auditorium 1, Haartmaninkatu 3,
on 7th November 2008, at 12 o'clock noon.*

Helsinki, 2008

Supervisors	Professor Timo Hakulinen Finnish Cancer Registry Helsinki, Finland and Department of Public Health University of Helsinki, Finland
	PhD Sirpa Heinävaara Radiation and Nuclear Safety Authority Helsinki, Finland
Reviewers	Professor Harry de Koning Erasmus University Rotterdam, The Netherlands
	Professor Esa Läärä Department of Mathematical Sciences University of Oulu, Finland
Official opponent	PhD Sue Moss Institute of Cancer Research London, United Kingdom

ISSN 0355-7979
ISBN 978-952-10-1383-6
ISBN 978-952-10-4852-4 (pdf)

Helsinki 2008
Helsinki University Print

Abstract

A population-based early detection program for breast cancer has been in progress in Finland since 1987. According to regulations during the study period 1987-2001, free of charge mammography screening was offered every second year to women aged 50-59 years. Recently, the screening service was decided to be extended to age group 50-69. However, the scope of the program is still frequently discussed in public and information about potential impacts of mass-screening practice changes on future breast cancer burden is required.

The aim of this doctoral thesis is to present methodologies for taking into account the mass-screening invitation information in breast cancer burden predictions, and to present alternative breast cancer incidence and mortality predictions up to 2012 based on scenarios of the future screening policy. The focus of this work is not on assessing the absolute efficacy but the effectiveness of mass-screening, and, by utilizing the data on invitations, on showing the estimated impacts of changes in an existing screening program on the short-term predictions.

The breast cancer mortality predictions are calculated using a model that combines incidence, cause-specific and other cause survival on individual level. The screening invitation data are incorporated into modeling of breast cancer incidence and survival by dividing the program into separate components (first and subsequent rounds and years within them, breaks, and post screening period) and defining a variable that gives the component of the screening program. The incidence is modeled using a Poisson regression approach and the breast cancer survival by applying a parametric mixture cure model, where the patient population is allowed to be a combination of cured and uncured patients. The patients' risk to die from other causes than breast cancer is allowed to differ from that of a corresponding general population group and to depend on age and follow-up time.

As a result, the effects of separate components of the screening program on incidence, proportion of cured and the survival of the uncured are quantified. According to the predictions, the impacts of policy changes, like extending the program from age group 50-59 to 50-69, are clearly visible on incidence while the effects on mortality in age group 40-74 are minor. Extending the screening service would increase the incidence of localized breast cancers but decrease the rates of non-localized breast cancer. There were no major differences between mortality predictions yielded by alternative future scenarios of the screening policy: Any policy change would have at the most a 3.0% reduction on overall breast cancer mortality compared to continuing the current practice in the near future.

Acknowledgements

This study was carried out at the Finnish Cancer Registry during years 2004-2008. It was funded by the Cancer Society of Finland, Doctoral Programs in Public Health at the Department of Public Health of the University of Helsinki, and MaDaMe research project of the Academy of Finland.

I am very grateful to Professor **Timo Hakulinen**, my principal supervisor and the Director of the Finnish Cancer Registry, for his expert guidance. He proposed me the topic, and along the way his ideas and strong theoretical skills have had a great impact on this work. He also seems to know all the research and people involved in cancer epidemiology, which has been a great help for me.

I also wish to thank my co-supervisor PhD **Sirpa Heinävaara**, not only for sharing her unique knowledge of the recent advances in the field, but also for the optimistic attitude and constructive grip on the thesis project. Her continuous support and friendship have been crucial for me during this process.

I want to express my sincere thanks to the co-authors Docent **Ahti Anttila**, PhD **Tytti Sarkeala**, Professor **Kaija Holli**, and M.Sc. **Hanna Virkkunen** for their collaboration.

The external reviewers of this thesis, Professor **Harry de Koning** from the Erasmus University in Rotterdam and Professor **Esa Läärä** from the Oulu University, are acknowledged for their valuable remarks.

My colleagues at the Mass Screening Registry, PhD **Tytti Sarkeala**, MD **Laura Kotaniemi-Talonen** and Docent **Nea Malila**, have offered me support, friendship and good company during these years – especially during the long lunch breaks. I wish our connection will last in a form or another despite the separate places of future employment. Many thanks also to the **Finnish Cancer Registry staff** members for providing the data, assistance with computers, hunting for literature, and help in numerous miscellaneous problems.

I would like to use the opportunity to thank my cousin and friend M.Sc. **Sari Koskinen-Kivilahti** for standing by me. Finally, my warmest thanks belong to my husband PhD **Juha Heikkinen**, who is both a brilliant husband and a brilliant Statistician. His infinite support, encouragement and help on all levels have kept me going on. Our children **Juulia**, **Jaakko** and **Saara** are an unfailing source of joy, love and energy for me.

Helsinki, September 2008
Johanna Seppänen

Contents

Abstract.....	3
Acknowledgements	4
Abbreviations	6
List of original publications.....	7
1. Introduction	8
2. Review of the literature	12
3. Aims of the thesis	14
4. Materials	15
4.1 Cancer registry data	15
4.2 Screening invitation data	15
4.3 Database in publications I-III	16
4.4 Database in publication IV	17
4.5 General restrictions	17
5. Methods	19
5.1 Adjustment for covariates (I-III)	19
5.1.1 Age and cohort	19
5.1.2 Screening program components	20
5.1.3 Region.....	23
5.2 Future screening policy scenarios (I, III).....	23
5.3 Incidence models (I-III)	24
5.4 Prediction model for incidence (I, III)	26
5.5 Parametric mixture model for breast cancer survival (III)	27
5.5.1 Variation in breast cancer survival	29
5.6 Survival from other causes than breast cancer (III)	31
5.7 Mortality predictions (III)	32
5.8 Cox model and relative survival (IV)	33
6. Results	34
6.1 Influence of mass-screening policy changes on incidence and mortality predictions for year 2012.....	34
6.2 Effects of separate screening program components on breast cancer incidence	36
6.3 Effects of separate screening program components on proportion of cured and life expectancy for uncured patients	38
6.4 Uncertainty in mortality predictions related to breast cancer survival	40
6.5 Cancer registry vs. clinical variables as predictors for breast cancer survival	41
7. Discussion.....	42
8. Conclusions	47
References	48

Abbreviations

BMI	Body mass index = $\text{weight(kg)} / \text{height(m)}^2$
CI	Confidence interval
DCIS	Ductal carcinoma <i>in situ</i> tumor, a non-invasive breast carcinoma
EUROCARE	European cancer registry–based project on the survival and care of patients with cancer
HRT	Hormone replacement therapy
IARC	International Agency for Research on Cancer (WHO’s cancer research institute)
RR	Relative risk

List of original publications

- I Seppänen J, Heinävaara S, Hakulinen T. Influence of alternative mammographic screening scenarios on breast cancer incidence predictions. *Cancer Causes and Control* 2006; 17:1135-1144.
- II Seppänen J, Heinävaara S, Anttila A, Sarkeala T, Virkkunen H, Hakulinen T. Effects of different phases of an invitational screening programme on breast cancer incidence. *International Journal of Cancer* 2006; 119:920-924.
- III Seppänen J, Heinävaara S, Hakulinen T. Predicting impacts of mass-screening policy changes on breast cancer mortality. *Statistics in Medicine* 2008; 27. Published Online 9.7.2008.
- IV Seppänen J, Heinävaara S, Holli K, Hakulinen T. Comparison of cancer registry and clinical data as predictors for breast cancer survival. *Cancer Causes and Control* 2008. Published Online 28.8.2008.

These publications will be cited in the text by their Roman numerals.

These original publications have been reprinted with kind permission of the copyright holders Springer Science and Business Media (I, IV) and John Wiley & Sons Ltd. (II, III).

1. Introduction

Breast cancer is the most common cancer among women in Europe and also world wide in developed countries [1, 2]. In Finland, breast cancer was responsible for 31.1% of all new cancer cases and 16.8% of cancer deaths among women in year 2006 (www.cancerregistry.fi). The number of new cases has been strikingly increasing during the past decades, and even if the age-specific rates remained constant, the ageing of the Finnish population will cause continuous increase of new cases in the future. However, due to improved treatments and an effective early detection program [3-7], the recent breast cancer mortality shows a constant trend, see Figure 1.

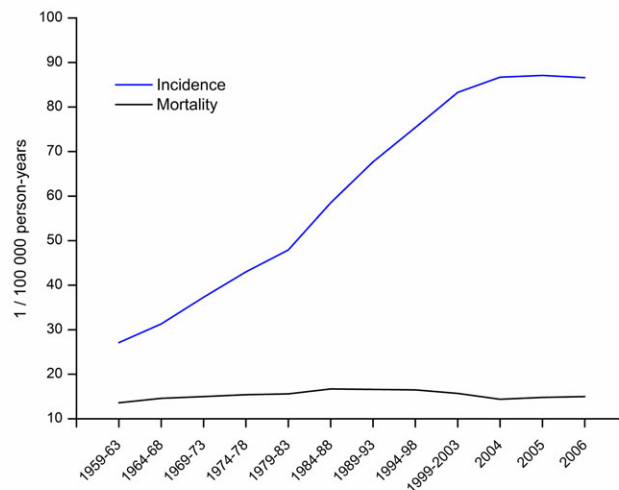


Figure 1: Age-adjusted (world standard) breast cancer incidence and mortality rates for Finnish women in time period 1959-2006. Data are from the Finnish Cancer Registry [www.cancerregistry.fi].

Breast cancer is one of the cancer sites subjected to the most extensive research. Well confirmed risk factors for breast cancer are reproductive factors (high age at first childbearing, early menarche, late menopause, few children, short breastfeeding period), use of HRT, genetic susceptibility, ionizing radiation, high breast density, and previous benign breast diseases [8]. There is also some evidence that some lifestyle factors such as low pre-menopausal or high postmenopausal BMI, high consumption of alcohol and low physical activity would increase the risk of breast cancer, but, except alcohol, their association appears to be weaker than that of the reproductive factors. Therefore, as many of the known risk factors are biological in nature or otherwise not relevant to be controlled with primary prevention in public health care, there is a distinct need for early detection programs.

Mammographic screening aims to find tumors in an early, detectable preclinical phase, and the core concept is to bring the diagnosis and start of treatment earlier in time when tumors are still small, localized and more responsive to the treatments. Since earlier diagnosis may improve the prognosis, and necessarily not so heavy, long and expensive treatments are needed, the early detection service is considered to be beneficial both for the patients and for the health care system. On the other hand, the mass-screening service is a substantial financial load for the Finnish municipalities who bear the responsibility for conducting the screening. Other disadvantages of screening are over-diagnosis, increased anxiety for screened women, and the exposure to extra radiation. Since the municipalities aim to lead evidence-based screening policy, the balance between pros and cons of mammography screening service is a continuous topic of public discussions.

In Finland, a population-based early detection program for breast cancer was introduced in 1987. It started gradually, but covered the whole country already in early 1990s. During the study period 1987-2001, there was a bylaw on public health regulating the invitation procedure by stating that organized, free of charge mammography screening should be offered every second year to women aged 50-59 years, while screening of 60-69 years old remained optional. The Finnish invitational screening program for breast cancer in ages 50-59 is reported to be effective [3-7], but information about potential impacts of extending to older and/or younger age groups or cutting the mass-screening practice on future breast cancer burden is required. This work is motivated by the lack of proper methodology for such a policy change assessment.

The measures of breast cancer burden are connected with each other; when we know the incidence and survival of the disease, we can calculate the prevalence and mortality. Publications I and II included in this thesis deal with estimating and predicting the incidence. The third paper (III) concentrates on estimating the survival and, by utilizing the results presented for incidence in paper I, to predict the breast cancer mortality. Based on different future scenarios of mass-screening policy, alternative short-term predictions for incidence (I) and mortality (III) for year 2012 are presented. The methodology is mostly developed in publications I and III, while publication II concentrates on quantifying the effects of separate components of the mass-screening program on breast cancer incidence. Paper (IV) provides important back-up for the predictions by studying the validity of used predictors. It focuses on finding out how important is the role of screen-detection and clinical tumor characteristics in addition to cancer registry information in describing the breast cancer survival. As the only covariates used in incidence and mortality prediction models in addition to screening invitation are the ones that are available in the cancer registry, namely age, stage and calendar time, it is crucial to know how powerful they are and how would the predictions change if we had all the relevant information of all the breast cancer cases in the population.

The term ‘efficacy’ has been defined by Last [9] as “the extent to which a specific intervention, procedure, regimen or service produces a beneficial result under ideal circumstances”, contrasting with the closely related term ‘effectiveness’ defined by Last as “a measure of the extent to which a specific intervention, procedure, regimen or service, when deployed in the field in routine circumstances, does what it is intended to do for a specific population”. In practice, studies to evaluate mass-screening for breast cancer have rarely assessed efficacy or effectiveness as defined, but rather a mixture of the two, depending on the design and other circumstances of the study [10]. The focus of this work is on assessing the effectiveness of mass-screening, and, by utilizing the data on invitations, on showing the estimated impacts of changes in an existing screening program on the short-term predictions. Given the screening process (2-year screening interval, attendance rate, rate of opportunistic screening, specificity, sensitivity etc.) what would happen to the breast cancer incidence and mortality in Finland if we extended the invitations to older or/and younger age groups or stopped the program entirely?

The backbone of the policy change assessments (I-III) is the novel screening invitation dataset, utilized for the first time in this thesis. It is a large data matrix that includes year- and age-specific information whether women in each municipality were invited or not to have a free mammography test during time period 1987-2001. First, this database makes it possible to take into account the real patterns of screening that have been going on in municipalities; we do not have to content with assuming that every woman was screened regularly every second year as stated by the official recommendations. It can be expected that modeling based on detailed municipal invitation data will result in more accurate predictions than those based on country level ignoring deviations. Secondly, data on year- and age-level enables adequate management of the two-year mass-screening cycle; the years within a screening round can be separated from each other, leading to a very specific manner of taking the program into pieces. The programming of the future screening policy scenarios is facilitated by the possibility to combine these pieces to correspond to the desired policy.

The most important indicator for the effectiveness of a mass-screening program is the cause-specific mortality. Screening activities without a decreasing impact on mortality are poorly justifiable. Even so, the incidence of breast cancer is a more acute indicator than mortality with respect to required health care resources. There is not much specific information about how the separate components (first and subsequent rounds and years within them, breaks, and post screening period) of an invitational mass-screening program differ from each other in terms of influencing the incidence. It is also reasonable to assume that different program components’ impact on breast cancer incidence is dependent on women’s age and stage of cancer at diagnosis, and statistical models should incorporate this heterogeneity.

As treatments get better and curability of breast cancer has become reality, it is reasonable to assume that part of the patients can be considered to be healthy survivors. This proportion of the patient population is not at risk of dying from breast cancer and can be assumed to be cured of the disease in a statistical sense. Thus, cancer survival analysis, which had focused almost exclusively on time to death, should also focus on estimating the proportion of long-term survivors among the patients. Following Heinävaara and Hakulinen [11] and De Angelis *et al.* [12], the breast cancer survival is modeled by applying a parametric mixture model where the patient population is allowed to be a mixture of two subpopulations with distinct risk of dying: those patients that are cured and those that are bound to die of the disease. This approach provides a way of modeling the hazard of fatal cases and the proportion of cured cases simultaneously. The role of prognostic factors, such as age and screening program component at diagnosis can be evaluated separately for these two patient groups. In addition, the proportion of cured patients can be allowed to vary by covariates. In this thesis, particular interest lies on the possible differences in cure fractions between separate components of the mass-screening program at diagnosis. This is the first time a cure model is fitted to Finnish breast cancer data.

To be able to die from breast cancer at a certain time point t , a patient must first stay alive until that. That is, not to die either from breast cancer or other causes before time t . For this reason we need to know the estimate for patients' survival from other causes than breast cancer in addition to cause-specific survival. It would be straightforward to assume that breast cancer patients are a random sample and their risk to die from other causes than breast cancer is the same as in corresponding general population. However, earlier results from stomach [13] and lung cancer [11] suggest that the risk could be significantly elevated and may also depend on covariates. The mortality model should therefore somehow account for the selection of patients with respect to general population mortality.

Since one aim of this study is to present methodologies for taking into account the mass-screening invitation information in breast cancer burden predictions, the presented models are essential parts of the results. So, despite the thesis is organized in separate sections for methods and results, the section '5. Methods' contains a considerable part of this work's results. This is a typical feature for studies that place themselves on the boundary between theoretical and applied fields of study, in this case on the borderline of statistics and epidemiology.

2. Review of the literature

The efficacy of a screening program is demonstrated by changes in disease-specific mortality, and there is a vast literature dealing with randomised studies to evaluate the efficacy of mammographic breast cancer screening programs world-wide, best summarized in the Handbook by International Agency for Research on Cancer [10]. In randomised controlled trials, screening women aged 50-69 has been shown to reduce mortality from breast cancer by approximately 25% [14]. The Finnish invitational screening program for breast cancer in ages 50-59 has been reported to reduce breast cancers in several studies from years 1999 to 2008 [3-7].

Studies to assess consequences of changing an existing mass-screening program are still rare. In their recent article, Sarkeala *et al.* conclude that extension of invitations to age group 60-69 would prevent breast cancer deaths among the elderly in Finland [15]. Moss *et al.* have estimated that starting the invitations already at age 40 years would yield a non-significant reduction in breast cancer mortality [16], and thus further follow-up is needed for any decisions. Impacts of mammographic screening policy alternatives on cost-effectiveness and quality of life have been studied by de Koning *et al.* [17]. An early study in 1981 concerning public health scenarios was made by Hakulinen *et al.* [18], who evaluated effects of hypothetical changes in men's smoking habits on lung cancer incidence predictions.

The focus of this study is on breast cancer burden predictions that are usually made by extrapolating the trend [19]. However, trend extrapolations may not be capable to produce accurate predictions when there are known factors, such as systematic screening, that might make abrupt changes in the temporal development of cancer incidence affecting potentially also the future mortality [20]. Such a sudden and unpredictable change has been recently visible in Finnish prostate cancer rates, which increased strongly in the beginning of this century mostly due to widely used PSA-testing [21].

Availability of a population-based registry database that includes information on screening invitations and/or results is not self-evident. Most of the earlier works on breast cancer incidence that include screening in the model assume regular screening activities as stated in the official guidelines [22, 23], or use estimated screening patterns [24]. The observed invitation schemes used in this work include also deviations from the bylaw on a municipality-level, which gives a more realistic picture of the true screening activities taking place in Finland than assuming fixed patterns.

The age-period-cohort model has been widely used for analyses of cancer incidence and mortality [22, 23, 25, 26]. Modeling of breast cancer incidence in papers I-III was

performed within a likelihood framework using the general age-period approach presented by Clayton and Schifflers [27], who also discussed the problem of non-identifiability in age-period-cohort models [28]. The concept of statistical cure and its maximum likelihood estimates were introduced by Boag in his article from 1949 [29]. Later on, De Angelis *et al.* [12] presented a mixture cure model for relative survival that incorporate background mortality using both exponential and Weibull survival distributions for the ‘uncured’ group. However, the shape parameter in the Weibull distribution was held constant and was not allowed to vary by covariates. Phillips *et al.* [13] introduced a parametric mixture model for relative survival where cancer patients’ hazard of dying from other causes did not need to be the same as in the corresponding group of the general population. This model has been extended to cause-specific survival [11, 30]. Also a non-mixture cure fraction model has been presented and extended to incorporate the background mortality [31].

Hakulinen *et al.* combined the predictions of incidence, general mortality and patients’ excess mortality together with population forecasts to predict cancer-specific mortality [32]. Verdecchia *et al.* developed a method called PIAMOD (Prevalence, Incidence, Analysis MODel) for the estimation of the future cancer burden and illustrated it for breast cancer [33]. Heinävaara *et al.* [11] have incorporated a parametric mixture cure model into a model that combines incidence, cause-specific and other cause survival on individual level to obtain short-time predictions for lung cancer prevalence and mortality. The current work extends the model of Heinävaara *et al.* by incorporating the screening invitation information through survival time for the uncured patients, proportion of cured, and shape and scale parameters of the underlying survival time distribution. The patients’ risk to die from other causes than breast cancer is allowed to differ from that of a corresponding general population group and to depend on age and follow-up time. In addition, hypothetical scenarios of future screening policy are programmed and used to calculate alternative incidence based breast cancer mortality predictions up to 2012, and an approach to assess the precision of the mortality predictions based on variation in the breast cancer mortality estimates is presented.

3. Aims of the thesis

The aims of this study were:

- To develop statistical methodology for taking historical municipality-specific schemes and alternative future scenarios of mass-screening into account when modeling and constructing predictions for incidence and mortality of breast cancer. **(I, III)**
- To quantify the effects of separate components of an invitational screening program on breast cancer incidence in Finland. **(I, II)**
- To produce predictions of future incidence and mortality of breast cancer in Finland based on alternative future scenarios of screening. **(I, III)**
- To compare breast cancer survival models including only cancer registry variables with models that additionally include screen-detection information and clinical tumor characteristics obtained from clinical data representative of the population. By comparing the models it is possible to evaluate the importance of the variables lacking from the Finnish Cancer Registry databases. **(IV)**

4. Materials

4.1 Cancer registry data

The breast cancer data were obtained from the Finnish Cancer Registry, which maintains a nation-wide database on all cancer cases diagnosed in residents of Finland since 1953. Almost 100% coverage of all breast cancer cases is ensured by compulsory, independent reporting from physicians, pathological laboratories, and hospitals, combined with supplementary death certificate information from Statistics Finland.

The stage of each breast cancer at diagnosis is coded at the Finnish Cancer Registry as follows: 0 = unreported, 1 = localized, 2 = positive regional lymph nodes, only regional spread, 3 = metastases, other than in regional lymph nodes, 4 = non-localized, unspecified. For the modeling, cases were classified to localized (stage 1), non-localized (stages 2, 3 and 4), and unreported (stage 0) cancers.

The cause of death information needed for cause-specific and other cause survival in publications III and IV has been separately evaluated for each patient at the Finnish Cancer Registry. In addition to the official cause of death received from Statistics Finland, this evaluation was based on the multiple subsequent clinical notifications sent to the Registry concerning each case. As result, the relationship between the cancer in question and the official cause of death is recorded with a special code.

4.2 Screening invitation data

The screening invitation data used in publications I-III were assembled from the Mass Screening Registry files where the registration of the Finnish screening program for breast cancer is centrally maintained. The Mass Screening Registry is part of the Finnish Cancer Registry. The database consists of year- and cohort-specific invitation information (invited/no invited) during the study period 1987-2001 in each of those Finnish municipalities that have made an arrangement with one of the ten screening centres of the Cancer Society of Finland. The age range is 40-74. In year 2001, these centres covered 267 out of 444 (60.1%) municipalities and 59.0% of 40-74 years old women in Finland. The mean compliance among 50-64 years old women was 90% at first screen and 93% at subsequent screens during the 90's [34], so the invitation status can be assumed to reflect the real screening activity very well.

Besides of conducting the mandatory screening of the age group 50-59, almost all municipalities (95%) offered some screening to age group 60-64, and 39% to age group 65-69. A total of 67 (25.1%) municipalities offered voluntarily screening to young women aged 40-49, and 19 (7.1%) to women aged 70-74. In general, minor deviations from the

bylaw, such as short, irregular breaks in the screening program and delayed starting were rather common among municipalities.

4.3 Database in publications I-III

To construct the database for publications I-III, the annual screening invitation schemes from period 1987-2001 were linked with corresponding female breast cancer data from the Finnish Cancer Registry and observed (1987-2002) and predicted (2003-2012) population count data received from Statistics Finland. The linking was done by municipality, calendar year and year of birth. Since every Finnish resident has a unique personal identification number, the linking and follow-up of individual records is simple and reliable.

Originally, the breast cancer data for publications I-III covered time period 1975-2001. However, the prediction base was restricted to 1987-2001 by practical reasons: the Finnish mass-screening program was implemented in 1987, so the incidence trend from earlier years was quite outdated. On the other hand, at the time of the data management process in spring 2004, the latest complete cancer data was from year 2001. Data from pre-screening years 1975-1986 were used to investigate the age-incidence and cohort-incidence dependences in publication I, see section 5.1.1. For breast cancer survival in publication III, the patients were followed up for death until the end of 2002. Only 24 (0.1%) patients were lost from follow-up, mainly due to emigration.

During 1987-2001 there were 907 326 women of age 40-74 in 267 Finnish municipalities in 13 074 cohorts included in the database (residence January 1 each year), and they contributed a total of 9 499 418 person-years to the study. Out of these women, 20 853 (2.3%) were diagnosed with a breast cancer during the period 1987-2001 and, as a consequence, 3 499 (16.8%) of them died from breast cancer before the end of the year 2002, see Table 1.

Table 1: The number of female breast cancer cases during 1987-2001 and subsequent breast cancer deaths during 1987-2002 in 267 Finnish municipalities (I-III). The bottom line displays the number of person years at risk in each age group.

	40-49		50-59		60-74		Total	
	<u>Cases</u>	<u>Deaths</u>	<u>Cases</u>	<u>Deaths</u>	<u>Cases</u>	<u>Deaths</u>	<u>Cases (%)</u>	<u>Deaths (%)</u>
Localized	2 510	199	4 492	269	4 748	402	11 750 (56.3)	870 (24.9)
Non-localized	1 902	604	2 357	672	3 038	1 176	7 297 (35.0)	2 452 (70.1)
Unreported	342	46	670	47	794	84	1 806 (8.7)	177 (5.0)
Total	4 754	849	7 519	988	8 580	1 662	20 853 (100)	3 499 (100)
Person-years (%)	3 363 603 (35.4)		2 697 449 (28.4)		3 438 366 (36.2)		9 499 418 (100)	

For mortality calculations in publication III, age- and calendar year –specific expected survival probabilities for the Finnish general population from Statistics Finland were included in the database.

4.4 Database in publication IV

For publication IV, clinical data including all female patients diagnosed with invasive breast cancer during years 1996-1997 in Tampere University Hospital Area (N=483) were linked with corresponding registry data from the Finnish Cancer Registry. The patients were followed up for death until the end of year 2005. Since there was no selection of included patients, the data could be considered as representative of the population. The clinical dataset was originally collected for the EURO CARE high resolution studies [35].

The study database included both cancer registry and clinical stage, see Table 2. In case of discrepancies between the stage and other tumor characteristics (for instance TMN-classification) in the notification sent to Finnish Cancer Registry, the Registry staff uses the relevant information given in all notifications concerning the current case to reconsider the given stage. In these cases the stage recorded in the registry might differ from the clinical one. The cancer registry stage was used in the main analyses, but additional analyses were conducted to examine the quality of the registry stage.

Table 2: The number of patients by stage as recorded in the cancer registry vs. stage in the clinical data (IV).

Cancer registry stage	Clinical stage			
	Localized	Non-localized	Unreported	Total
Localized	200	53	37	290
Non-localized	1	151	9	161
Unreported	16	10	6	32
Total	217	214	52	483

The screen-detected cases (41.9%) were defined as tumors detected by population-based mammography screening or by mammography control on own initiative. Unlike in papers I-III, this dichotomous variable (yes/no) tells about the performance of the screening test, not the invitation. During years 1996 and 1997, the city of Tampere screened women up to the age 68 [36], while the screening policies varied between the municipalities in surroundings. A total of 48.3% of the patients were from the city of Tampere, the rest coming altogether from 27 municipalities (residence January 1 in the year of diagnosis). The proportion of cases diagnosed as a consequence of a test taken on women's own initiative was unknown.

4.5 General restrictions

Ductal carcinoma *in situ* (DCIS) tumors were excluded from the analyses, except in paper I, where they fell in both the localized and unreported groups (3.7% of all cases). When

estimating breast cancer survival in publications III and IV, cases obtained only from death certificates or autopsy reports were excluded from the analyses (0.1% and 0.7% respectively). Women with multiple breast cancers (2.0% in III and 0% in IV) were included only once in survival analysis according to their first breast cancer.

5. Methods

5.1 Adjustment for covariates (I-III)

No analysis in cancer epidemiology can ignore the effect of age, and this holds especially for breast cancer even after restricting the analysis to women in their potential mammographic mass-screening ages. There has also been a distinct increase in breast cancer incidence during the last decades [37], so, the prediction base being 15 years, inclusion of calendar time is unavoidable. Calendar time is also needed to enable the extrapolation into the future. Moreover, in the absence of reliable registry information on cancer treatment, the calendar time adjusts to a certain extent for the changes in treatment practices. Since screening has – by definition – a different impact on the incidence of localized than non-localized breast cancers, and because survival from breast cancer is dependent from stage (Finnish Cancer Registry 2003-2005), separate incidence and survival models were built for localized and non-localized cases. The overall estimates for incidence were done by separate modeling of all cases combined group, which included also the cases with unreported stage, meanwhile overall mortality predictions were calculated as a combination of the stage-specific predictions. The other included explanatory factors were the component of screening program at the year of diagnosis and the university hospital region. The same predictors were chosen to explain both the breast cancer incidence and mortality. However, the roles of these factors turned out to differ from model to another. These issues are covered in detail in the current section 5.1, followed by presentation of the future scenarios in section 5.2 and model definitions in the rest of the chapter 5.

All the data management and analyses were carried out with SAS Releases 8 and 9 [38]. Graphs were produced with the Origin scientific graphing package, version 7.5.

5.1.1 Age and cohort

To exclude confounding by regular mass-screening, data from pre-screening period 1975-1986 were used to look for a numerical function for the dependency between age and incidence in paper I [39]. A polynomial describing the age-incidence curve was determined by first fitting a simple Poisson regression model including continuous calendar time and numerical age as covariates. Then, one higher degree age term was added to the model at a time as long as they significantly ($p < 0.05$) improved the fit. Comparisons between these nested models were done with likelihood ratio tests. As result, the age-incidence dependency was described with a polynomial of fourth degree both in localized and non-localized breast cancers. After controlling for calendar time and fourth degree polynomial age dependency, the cohort term was subjected to a similar procedure, but, based on likelihood ratio testing, there was no significant cohort effect left. This is, however, mostly due to the linear dependency among age, period and cohort, and does not imply that cohort itself has no effect on breast cancer incidence. For mortality predictions in paper III, the breast cancer incidence was modeled similarly to paper I.

In order to account for the possible changes in age distribution of Finnish 40-74 years old women during the period 1987-2012, age standardization was tried for both the observed and predicted breast cancer incidences in paper I. The age distribution in year 2001 in the middle of the period was chosen to be the standard. This had, however, no major influences on the incidence, and the crude rates were used in all publications.

In paper II, continuous age and cohort terms in addition to 5-year age group and calendar time were not needed. This was mainly due to restriction of the analysis to ages 50-74 instead of 40-74 and using 5-year age groups instead of 10-year. After excluding the youngest women and adjusting for screening the relationship within 5-year age categories turned out to be constant. In addition, the effect of age at initial invitation was examined by adding into the model a variable indicating whether the municipality has started the invitations of a cohort before the age of 55 (33% of cohorts) or thereafter (25.5%): It showed marginal significance ($0.05 < p < 0.1$) in the analyses of localized cancers and all cases combined, but since it had no essential impact on the results it was eventually left out from the model.

When modeling the breast cancer survival (III), the effect of numerical age at diagnosis was found to be linear on the log-scale, but categorical age in addition to numerical improved the fit markedly, so they were both included in the model. The categorized age defines the level of the group and the numerical age the change within the group. By adding the categorical age in the model, the effect of linear age was allowed to differ between the age categories.

An interaction term was needed in both incidence models (I, II) to describe the differences between age groups within a program component. Because the interaction was not wanted to take a numerical form, categorical age was used for that purpose. However, in the model for breast cancer survival, an interaction term between age and screening invitation component did not improve the fit.

5.1.2 Screening program components

During the observation period 1987-2001, the Finnish mass-screening service covered ages 50-59 as mandatory and the screening cycle was two years. To be able to incorporate the screening invitation data into statistical models, the screening program was divided into components. The division was based on an assumption that different parts of an invitational screening program have heterogeneous impacts on breast cancer incidence (I, II). Invitation to mammography, as well as in the majority of women the mammography test itself, takes place during the first year of the screening round. If screening has an impact on breast cancer incidence, the incidence cannot be assumed to be constant during the whole round, and therefore the two years within one round were coded separately. The first screening round, known as the prevalence round, was also separated from the subsequent ones; the increase in incidence was expected to be larger during the first round compared to the subsequent ones, mainly because of the detection of cases from the pool of prevalent undiagnosed breast

cancers [10]. In addition, the five immediate years after the program were separated from the rest of the follow-up time [22], and the observed irregular breaks were identified. There were originally nine components in the program:

- 1 = 1st screening round / 1st year
 - 2 = 1st screening round / 2nd year
 - 3 = Subsequent screening rounds / 1st year
 - 4 = Subsequent screening rounds / 2nd year
 - 5 = Up to 5 years after the last screening round
 - 6 = More than 5 years after the last round
 - 7 = Break
 - 8 = Screening round following 2-3 years after the break in the program / 1st year
 - 9 = Screening round following 2-3 years after the break in the program / 2nd year
- 0 = Not invited

The component 7 was defined as a break in a cohort's screening program. These are irregular years in between the program when the municipalities have deviated from the guidelines; the main point is that the cohort has been screened at least once and will be screened again in the future. After a break of more than four years between invitations the coding starts with component 1 again, assuming that any carry-over effects of the last screening would have disappeared. Originally, there were also separate codes for the first round following a 2-3 years break in the program, but the cases diagnosed during that component were so few (0.6% of all cases) that they were combined with the 'normal' subsequent rounds in the model-building process (I, II). The majority (87.5%) of these cases were localized.

It should be noted that, because of the time lag between invitation to screening test and the evaluation of the outcome, cancers diagnosed during the breaks in this setting do not meet the criteria of an 'interval cancer', defined as cancers diagnosed clinically between the screening rounds among those with negative results at screening [10].

The component 0, not invited, includes all the women up to the point of first invitation. This component served as a baseline. In most of the municipalities, these were women below 50 (some municipalities started the screening already at the age of 40) throughout the whole period 1987-2001, or above 59 at the time when screening started. Because the Finnish program was introduced gradually, and because the deviations from the bylaw, there were also cohorts that never got invited, or that were for some reason invited for the first time in older ages than 50. Every cohort was coded separately according to the municipality's invitation scheme, so all women living in a municipality and born in the same year had the same invitation pattern.

In the survival model for mortality calculations, the breast cancer deaths occurring during the follow-up were connected with the screening program component that took place during

the year of diagnosis (III). When estimating the parameters of the cure model, the reduction of the number of parameters became unavoidable in order to provide stable estimates for each subgroup. Components with similar estimates were combined on the basis of preliminary cause-specific survival analyses. Given the stage, it turned out that the prevalence screening round was quite similar to the subsequent ones in terms of breast cancer survival, but the difference between the years within a round was very clearly pronounced also in explaining the survival from breast cancer. Consequently, for modeling the survival time, the rounds were combined but the separation of the first and second years remained. The first five years after the program were also combined with the second years. To avoid confusion, the combined components were marked with capital letters:

A = 1st years
 B = 2nd years and up to 5 years after
 C = Break
 D = More than 5 years after

N = Not invited

Moreover, for model of the cure proportion P (section 5.5) only the first years were separated from the rest of the program marked with O. Again, this combination was done in order to ensure the convergence of the model and stable estimates. Table 3 displays the correspondence between the screening program components in incidence (I, II) and survival models (III). Figure 2 demonstrates the connection between the program component at diagnosis and subsequent breast cancer deaths occurring during the follow-up years.

Table 3: The original screening program components, and corresponding components in incidence (I, II) and survival (III) models. P is proportion of cured patients in the overall breast cancer survival model.

Original data	Incidence model	Survival time model (uncured)	Model of P
1	1 → A	A	A
2	2 → B	B → O	O
3	3 → A	C → O	
4	4 → B	D → O	
5	5 → B		
6	6 → D		
7	7 → C		
8 → 3			
9 → 4			
0	0 → N	N	N

given municipality, calendar year and age at diagnosis to alter from scenario to another. In all scenarios, it was assumed that municipalities will strictly follow the guidelines, that is, there will be no deviations or breaks in the program. The scenarios were the following:

- Scenario A: continuing to fulfill the current bylaw requirements of inviting 50-59 years old women every second year,
- Scenario B: expanding screening service from 50-59 to 50-65 years old,
- Scenario C: expanding screening service from 50-59 to 50-69 years old,
- Scenario D: expanding screening service from 50-59 to 40-65 years old women,
- Scenario E: stopping the screening service entirely.

Scenarios B and C are realistic and were actually happening in several Finnish municipalities during 1987-2001. Due to lack of evidence [16], relatively low incidence rates and breast tissue and tumor characteristics, there were no plans to expand the mass-screening to women aged below 50 years in Finland, but scenario D was still thought to be of interest. Scenario E reflects the situation after stopping the program: those women that have been screened just before 2005 were first coded to be in the component 5 'up to 5 years after the last screening round', followed by component 6 'more than 5 years after the last round', see section 5.1.2. Those who were to become 50 in year 2005 or thereafter were coded as 'not invited'. Scenario E was omitted from mortality predictions (III), since the observed data on breast cancer deaths during the post screening period were too sparse in women under 60 years of age. When predicting incidence based on scenario E, the lack of post screening data on age group 50-59 was compensated by assuming the relative risk estimate to be 1 (no effect) in the calculations.

All changes in the screening practice were programmed to take place in the beginning of 2005. The transitional period 2002-2004 from current practice to the future scenarios was programmed to follow the bylaw without any deviations. Because the observed schemes differed to some extent from one municipality to another, the major variations were accounted for when programming the transition period. For instance, if a municipality had invited women below 50 just before 2001, they were not coded to have their first year of the first round as they become 50, but the first year of a subsequent round etc.

In mortality predictions in publication III, scenarios A, C and D were used and renamed with Roman numerals as scenario I, II and III, correspondingly.

5.3 Incidence models (I-III)

The modeling of breast cancer incidence in papers I-III was performed within a likelihood framework using the general age-period approach presented by Clayton and Schifflers [27, 28]. The analyses were carried out using Poisson regression with a logarithmic link function. The models were built separately for localized, non-localized and all breast cancer cases combined, the latter group including also the cases with unreported stage.

In papers I and III, the stage-specific expected incidence EL_{mya} in municipality m ($m=1, \dots, 267$) in calendar year y ($y=1987, \dots, 2001$) for women of age a ($a=40, \dots, 74$) is expressed as

$$\log(EL_{mya}) = \alpha^{(I)} + \beta^{(I)}y + \gamma_1^{(I)}a + \gamma_2^{(I)}a^2 + \gamma_3^{(I)}a^3 + \gamma_4^{(I)}a^4 + \rho_{r(m)}^{(I)} + \delta_{s(m,y,a),c(a)}^{(I)} \quad (1)$$

Here $r(m)$ is the university hospital region r ($r=1, \dots, 5$) to which municipality m belongs, $s(m,y,a)$ is the screening invitation component s ($s=0, \dots, 7$) in municipality m in calendar year y for women of age a , and $c(a)$ is the age category c ($c=40-49, 50-59, 60-74$) that includes age a . The term $\delta_{s(m,y,a),c(a)}^{(I)}$ is an interaction between screening invitation status and categorical age. Calendar year and age were treated as numerical variables, and they were re-scaled for the analyses so that the origin of time was fixed to 2003 and age to 40. The superscript I stands for incidence to avoid later confusion with corresponding parameters in the survival model (see section 5.5). Figure 3 shows the fit of the incidence model (1).

In paper II, the breast cancer incidence was modeled similarly as in papers I and III, except that the age polynomial was replaced with categorical age (50-54, 55-59, 60-64, 65-69 and 70-74), see section 5.1.1. To demonstrate the hypothetical situation where the municipalities had followed the official recommendations conscientiously, cumulative incidences for fictional regular screening programs from 50 up to 59 and 69 years were calculated. The term ‘regular’ indicates that women are invited every second year, and no breaks or other deviations from the official recommendations take place. Calculations were done using the age- and program component -specific risk ratio estimates obtained as a result from modeling the breast cancer incidence in the observed data from years 1987-2001 including irregularities in the invitations. The estimated incidence of the non-invited women was used as the baseline level. The cumulative incidence (CI_a) at numerical age a ($a=50, \dots, 74$) in a regular screening program was calculated as

$$CI_a = \sum_{i=50}^{a-1} RR_i I_{B,i} , \quad (2)$$

where RR_i is the relative risk of breast cancer at age i compared to non-invited women at the same age, and $I_{B,i}$ is the breast cancer incidence estimate at age i among the non-invited women. The relative risk corresponds to the regular screening program component in each age, for example, in the regular program from 50 to 59 the women would be invited at the age of 50 (first year of the first round), 52, 54, 56 and 58 (first years of the subsequent rounds).

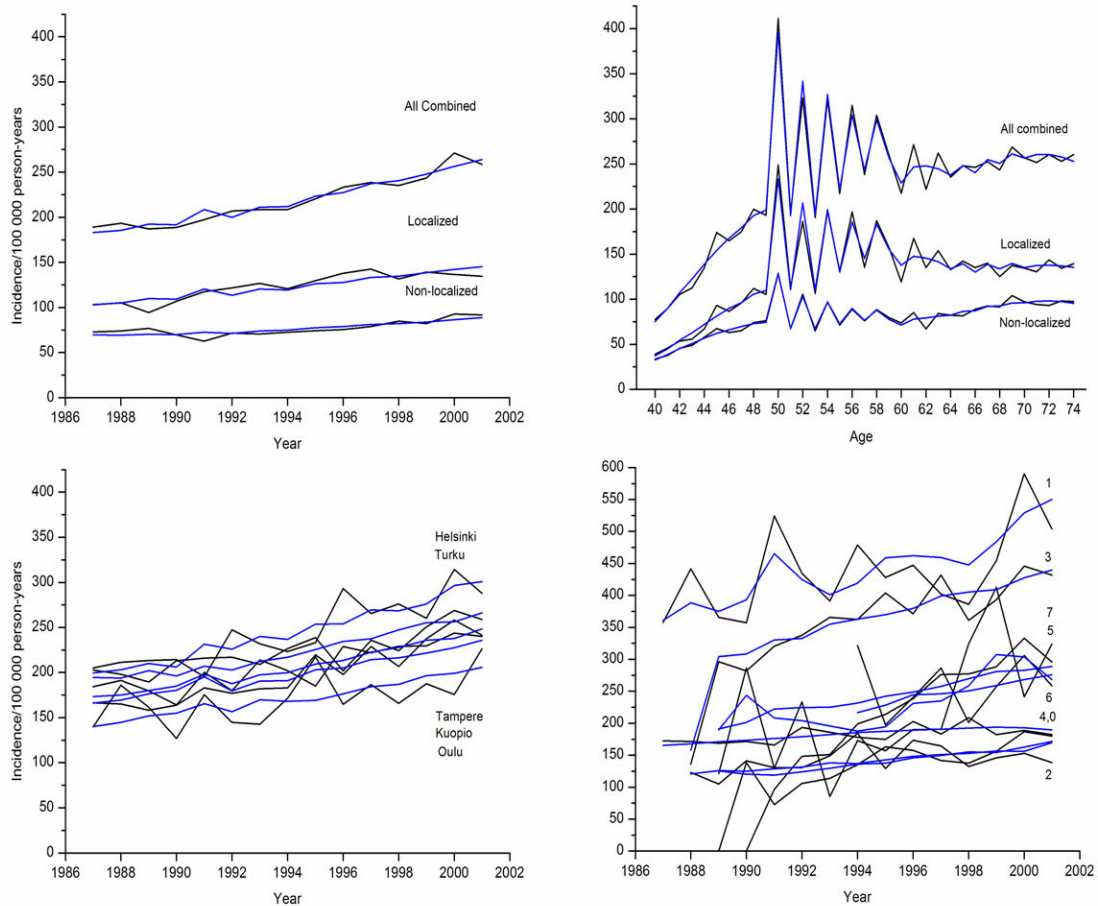


Figure 3: Observed (black) and model-based (blue) crude incidence curves by stage, age at diagnosis, university hospital region and screening program component. The fit by region and screening component is shown only for all stages combined. The program components are: 1 = 1st screening round / 1st year, 2 = 1st screening round / 2nd year, 3 = Subsequent screening rounds / 1st year, 4 = Subsequent screening rounds / 2nd year, 5 = Up to 5 years after the last screening round, 6 = More than 5 years after the last round, 7 = Break, 0 = Not invited.

5.4 Prediction model for incidence (I, III)

When calculating the predictions, the calendar time is the only term in the model that is extrapolated beyond the range of observations. Accordingly, there were two alternative approaches to extrapolate the incidence to the target year 2012. A natural first choice was to

use the same exponential model (1) for predictions as for modeling the observed incidence. The problem with that model is, however, that it can lead to unrealistic exponential growth in incidence of a cancer site with an increasing trend like in breast cancer. In the second approach, the exponential growth was leveled off by replacing the exponential relationship between incidence and calendar time with a linear one [19]. In short time predictions like these, the two approaches will produce predictions that differ only slightly from each other ($\exp(\beta^{(I)}y) \approx 1 + \beta^{(I)}y$), but when the target year is further in the future, leveling off the exponential growth with calendar time may become crucial. The second model with linear calendar time was chosen for making the future predictions based on the alternative scenarios.

The maximum likelihood estimates obtained as a result of modeling the breast cancer incidence (1) together with different scenarios of future screening policy (see section 5.2) were then plugged in to the prediction model to obtain alternative predictions for breast cancer incidence rates up to 2012.

The overall year-specific incidence rate predictions \hat{I}_y were derived by first calculating the predicted municipality-year-age-specific cancer case numbers $\hat{c}_{mya} = \hat{I}_{mya} \hat{n}_{mya}$, where \hat{n}_{mya} is the corresponding predicted number of person-years. These predicted numbers of cases and person-years were then summed over all municipalities and ages to obtain the respective year-specific figures: $\hat{c}_y = \sum_m \sum_a \hat{c}_{may}$ and $\hat{n}_y = \sum_m \sum_a \hat{n}_{may}$. Finally, the predicted year-specific incidence rates were calculated as $\hat{I}_y = \hat{c}_y / \hat{n}_y$.

5.5 Parametric mixture model for breast cancer survival (III)

Following Heinävaara and Hakulinen [11] and De Angelis *et al.* [12], the cause-specific survival needed for calculating mortality predictions was modeled with a parametric mixture model where the patient population was allowed to be a mixture of two subpopulations with distinct risk of dying: those patients that are cured and those that are bound to die of the disease (III). This approach provides a way of modeling the hazard of fatal cases and the proportion of cured cases simultaneously. Furthermore, the prognostic factors can be estimated separately for these two patient groups. The modeling was done separately for localized and non-localized breast cancer cases.

Let us assume that a proportion P of patients is cured and hence has a survival from breast cancer equal to 1; the remaining proportion $1-P$ represents those uncured. The mixture model for the overall breast cancer survival S_C can then be written as

$$S_C(t) = P + (1 - P)S_{C,1-P}(t), \quad (3)$$

where t is the survival time from breast cancer and $S_{C,1-P}$ is the breast cancer survival function for the uncured population. The proportion of cured P was allowed to depend on age category as in (1) and on a 3-class screening invitation status, see section 5.1.2 and Table 3.

The proportion of statistically cured patients P can be described as the asymptote of the overall survival curve S_C and the point of statistical cure for the cured population as the point in follow-up time when S_C reaches the level of P , that is, the point in time when S_C approaches its asymptote. This is illustrated in Figure 4. A proportion of cured patients can be observed when those remaining alive can be considered to be healthy survivors with respect to breast cancer; the situation is essentially equivalent to the interval specific relative survival being equal to one or excess mortality rate to zero.

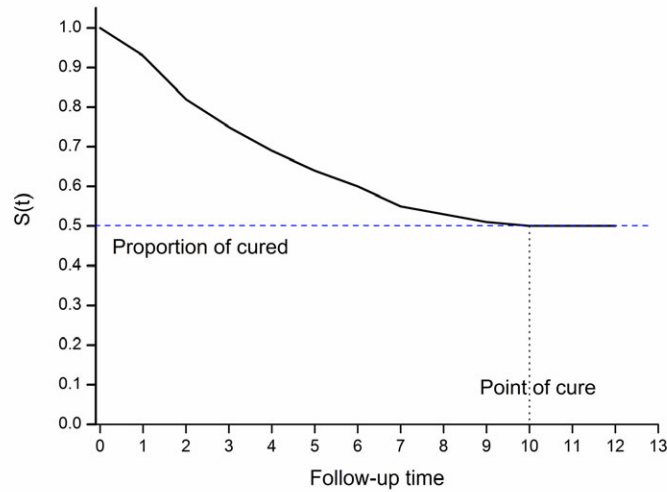


Figure 4: The proportion of statistically cured patients can be defined as the asymptote of the survival curve, and the point of statistical cure for the cured population as the point in follow-up time when the curve reaches the asymptote.

The expected survival times for the uncured patients were modeled as a function of year y , age a , age category $c(a)$ defined as in (1), and a combined screening invitation component $q(m, y, a)$ in municipality m in calendar year y for women of age a at diagnosis (see section 5.1.2 and Table 3):

$$\log(E(t)) = \alpha^{(S)} + \beta^{(S)}y + \gamma^{(S)}a + \sigma_{c(a)}^{(S)} + \delta_{q(m, y, a)}^{(S)}. \quad (4)$$

The superscript S stands for survival.

The log of the likelihood of the overall survival model (3) was maximized using the iterative Gauss-Newton method. It was assumed that only uncured patients can die from breast cancer. The computing was performed using the NLIN procedure in SAS [38].

Selection of the probability distribution for survival time and the model-building were based both on likelihood ratio tests and graphical diagnostics including examination of (log)negative-log plots and Cox-Snell residuals. The general principle was to make the model as simple and robust as possible. Since the primary aim was to provide reliable extrapolations, the fit of model-based number of deaths curve to the observed one, especially towards the end of the observation period, was the most important criterion for the goodness-of-fit of the survival model. As a result, the two parameter gamma distribution was chosen both on grounds of the best fit in both stage categories, and the flexibility it has in the shape of the function compared to other candidates exponential, Weibull, and lognormal. To allow that flexibility, the shape and scale parameters were expressed as functions of covariates. Figure 5 shows the fit of the survival model (3) by covariates.

5.5.1 Variation in breast cancer survival

To estimate the uncertainty related with breast cancer survival (3), and especially with its components life expectancy for uncured (4) and proportion of cured patients, a likelihood-based 95% confidence region was constructed

$$2(L_{\max} - L) \leq \chi^2_{(df=2, \alpha=0.05)} , \quad (5)$$

where L_{\max} is the maximum of the log-likelihood of the overall survival model and L is a log-likelihood obtained by altering the values of expected survival time and proportion of cured. When calculating L , the shape and scale parameters were not changed but held fixed as their maximum likelihood estimates.

After constructing the confidence region, different points were systematically chosen from the border of the region, and mortality predictions were re-calculated using corresponding values of life expectancy (LE) and P . The 95% confidence bounds for breast cancer mortality predictions related with uncertainty in breast cancer survival estimates were defined as the annual minimum and maximum of these predictions.

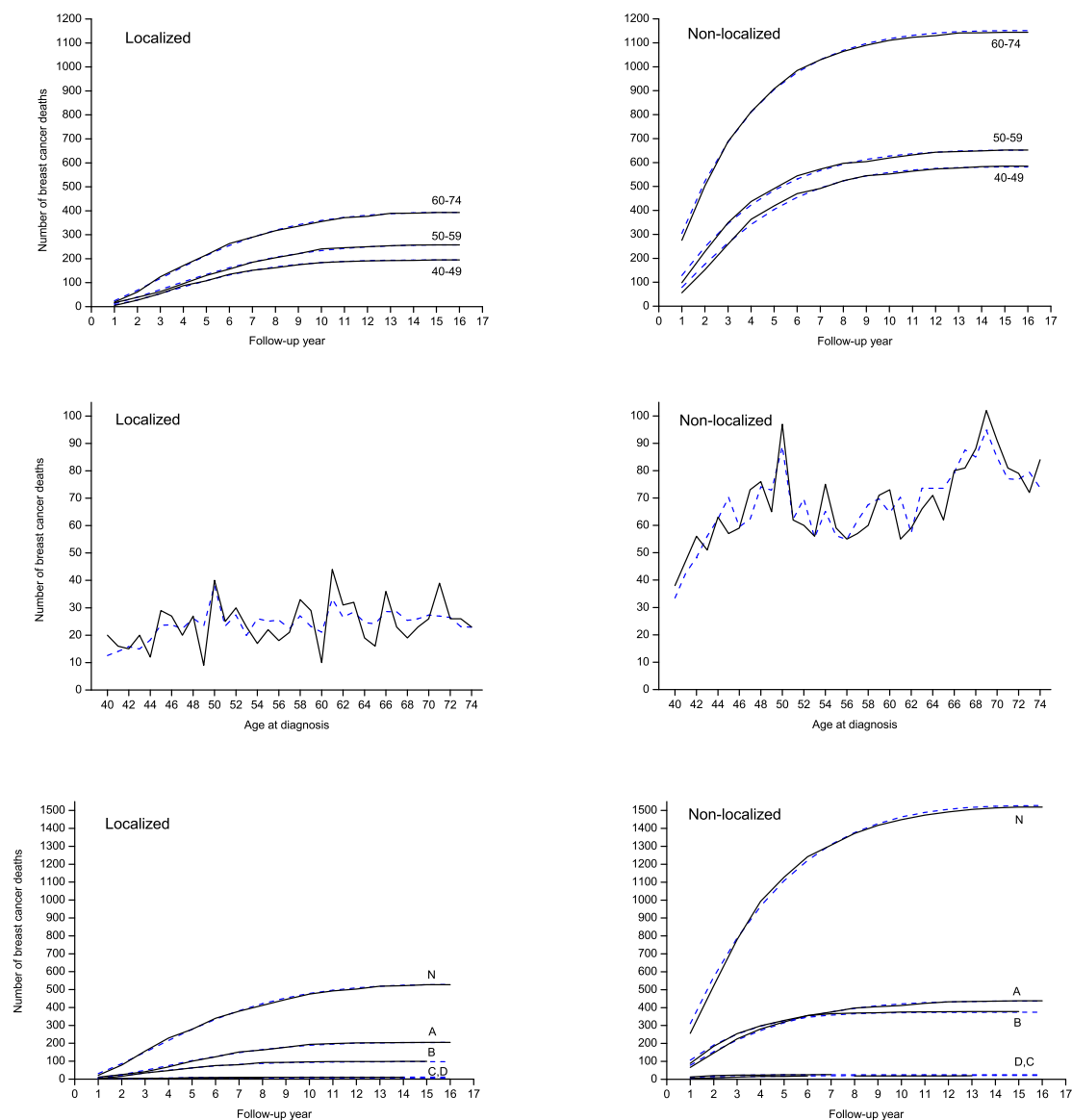


Figure 5: Observed (black solid line) vs. model-based (blue dashed line) numbers of breast cancer deaths during the observation period 1987-2001 plotted by grouped age, numerical age, and combined screening program component. All the covariates are measured at diagnosis. The combined program components are: A = 1st years, B = 2nd years and up to 5 years after, C = Break, D = More than 5 years after, N = Not invited.

5.6 Survival from other causes than breast cancer (III)

When predicting breast cancer mortality (III), we need to estimate the patient's other cause mortality as well; to be able to die from breast cancer at a certain time point t , a patient must first stay alive until that. That is, not to die either from breast cancer or other causes before time t .

The cancer patients' other cause mortality hazard could be derived directly from the general population life-tables. However, previous results from other cancer sites suggest that the cancer patients' risk to die from other causes is not necessarily the same as in corresponding general population [11, 13]. The relationship between patients' and general population's hazards might also not be constant but depend on covariates like stage of the tumor, age, and time passed from diagnosis. To account for that, the ratio between patients' hazard $h^{(O)}$ and general population's hazard h^* was used as a correction coefficient for the general population other cause mortality hazard. The general Finnish population was matched to the patient population by calendar year and numerical age. To assess the proportionality between hazards, the ratio $r = h^{(O)}/h^*$ was modeled as a function of follow-up period. The analysis was done using Poisson regression with a logarithmic link function and it was stratified by stage and age group. Figure 6 displays the observed and estimated hazard ratios.

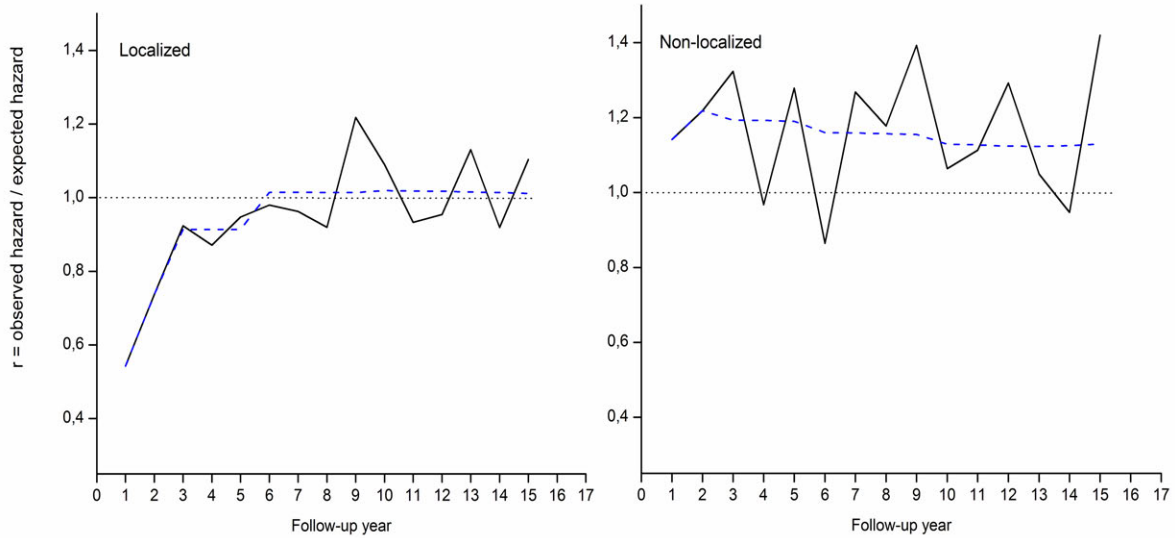


Figure 6: Observed (black solid line) vs. model-based (blue dashed line) ratios between patients' and general population's hazard $r = h^{(O)}/h^*$ for localized and non-localized cases.

Survival from other causes than breast cancer at follow-up time t (in years) for women diagnosed in calendar year y at age a was then expressed as

$$S_{ya}^{(O)}(t) = \exp \left[- \sum_{j=1}^t h^*((a-1) + j, (y-1) + j) \cdot r(c(a), j) \right], \quad (6)$$

where $h^*(u, v)$ is the hazard to die at age u in calendar year v in the general population, and $r(c, p)$ is the model-based ratio during follow-up period p ($p = 1, 2, 3-5, 6-9$, and 10 or more years after diagnosis) for women whose age at diagnosis is included in age group c ($c = 40-49, 50-59, 60-74$).

5.7 Mortality predictions (III)

The breast cancer mortality was calculated as a product of the observed (1987-2001) or predicted (2002-2012) number of new cases (I), predicted survival from breast cancer, and survival from other causes (III). This was possible because the log-likelihoods of survival from breast cancer and survival from other causes are additive. The combination of incidence and patient-level survival was done by assuming that the survival is the same among a group of women diagnosed with a breast cancer of a certain stage, age and year, and during the same component of screening.

If we denote the survival from breast cancer in a certain stage with $S^{(C)}$, survival from other causes with $S^{(O)}$, and the hazard to die from breast cancer with $h^{(C)}$, the expected number of breast cancer deaths D at follow-up time t (in months) in municipality m among women diagnosed in calendar year y in month z ($z=1, \dots, 12$) at age a is

$$D(t; m, y, z, a) = \frac{1}{12} B(m, y, a) \cdot S_{mya}^{(C)}(t; m, y, z, a) \cdot S_{ya}^{(O)}(t; y, z, a) \cdot h_{mya}^{(C)}(t; m, y, z, a), \quad (7)$$

where $B(m, y, a)$ is the number of cancer cases of the same stage observed (1987-2001) or predicted (2002-2012) to occur in municipality m in calendar year y for women of age a . The predicted number of cases as well as the screening component in $S^{(C)}$ alters depending on the future scenario used as the basis of incidence predictions, as described in section 5.2. The yearly number of new cases is assumed to be evenly distributed between months, so the number of cases in each month is $B/12$. The diagnosis was programmed to take place in the middle of the diagnosis month z . The hazard of dying from other causes than breast cancer was assumed to be constant during each calendar year.

The overall predictions were calculated as a combination of the stage-specific mortality predictions. However, there were 8.7% of the cases with an unreported stage in the database (see Table 1 in section 4.3) which had to be accounted for. In reality, the

unreported group consists of localized and non-localized cases in an unknown ratio. This ratio was estimated using relative survival, assuming that $x\%$ of the unreported cases would behave as localized and the rest as non-localized with respect to the relative survival. These proportions were then used to obtain stage-specific coefficients for calculating the overall number of deaths, and further the mortality.

5.8 Cox model and relative survival (IV)

Comparison of breast cancer survival models including only the cancer registry variables age and stage to models that include also screen-detection information and clinical tumor characteristics was carried out with the Cox regression for cause-specific survival. In order to examine the robustness of the results, the same models were subjected to relative survival analysis to estimate the excess risk associated with breast cancer.

Three hierarchical models were built to study the role of screen-detection and clinical tumor characteristics in addition to registry data in explaining the breast cancer survival: Model 1 included cancer registry variables stage of cancer (localized/non-localized) and age (40-49, 50-59, 60-74) at diagnosis. In addition to Model 1, Model 2 included information on whether the tumor was detected by screening (yes/no). Model 3 included same variables as Model 2 and the following clinical variables: size of tumor (range 3-130 mm), grade (1-3), total number of examined lymph nodes (0-29), number of metastatic lymph nodes (0-23), and estrogen receptor status (+/-). When modeling the relative survival, the follow-up time (1, 2, 3, 4, 5-10 years) was added as a factor in the models.

The relative survival ratio, defined as the ratio between observed survival in the patient group and expected survival in a comparable group from the general population, was modeled using a generalized linear model with a Poisson error structure [40]. The assumption of proportional hazards in Cox model was checked by adding interaction terms between follow-up time and other variables one at the time in relative survival Model 1. According to likelihood ratio testing, all the interaction terms were non-significant ($p > 0.05$), indicating no need for a departure from the assumption of a constant ratio of the hazards over time.

6. Results

6.1 Influence of mass-screening policy changes on incidence and mortality predictions for year 2012

The alternative incidence and mortality predictions for the target year 2012 are displayed in Figure 7. In general, the impacts of policy changes are clearly visible in the predictions for breast cancer incidence, while the effects on mortality are very minor. Furthermore, the changes influence more localized than non-localized incidence rates, the effects on breast cancer mortality being the opposite.

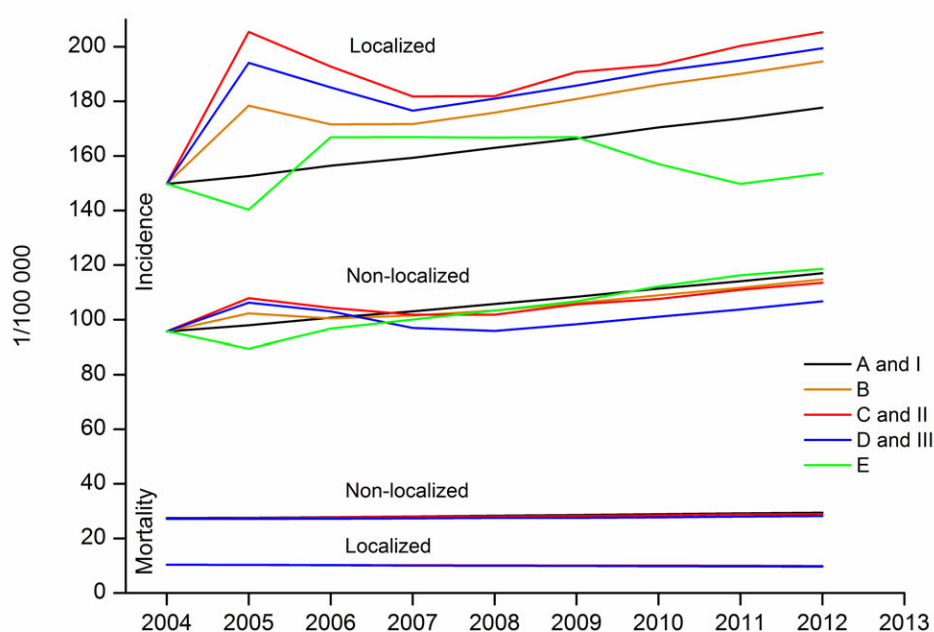


Figure 7: Predicted crude breast cancer incidence and mortality rates based on alternative screening policy scenarios. In scenario A (and I) the current practice of inviting 50-59 years old women every second year is continued, in scenario B the screening service is expanded from 50-59 to 50-65, in scenario C (and II) to 50-69, and in scenario D (and III) to 40-65 years old women. In scenario E the screening service is stopped entirely. Changes in screening practice have been programmed to take place in year 2005.

In localized cases, expansions of the screening program increase the predicted incidence rates compared to continuing the current practice regularly (scenario A) (Figure 7). The biggest increase (13.5%) is introduced by extending the service from current practice to ten years older women (scenario C). Stopping the program (scenario E) first causes a decline

due to the small RR estimates for subsequent rounds/2nd years in age groups 50-59 (Table 5), then an increase due to the 'up to 5 years after' estimate of 1.88 in the age group of 55-59 (RR was assumed as 1 in calculations for age group 50-54, see page 22). After year 2009 the 'more than 5 years after' estimates dominate until year 2011 when the code non-invited (RR=1) starts to take over. The impacts of policy changes on non-localized breast cancer incidence are smaller but protective; the biggest decrease (9.7%) would be achieved by extending to age group 40-65; the same applies to mortality, where the corresponding policy change would introduce a 4.5% reduction in short term predictions for non-localized cases compared to continuing the current practice.

Table 4 shows that the overall breast cancer incidence is strongly increasing according to all scenarios compared to year 2001, the biggest increase suggested by scenario II. Despite the increasing number of new cancer cases, the predicted mortality figures remain quite stable. The increasing effect of mass-screening on incidence is visible when comparing the age group -specific rates between scenarios. For instance, when extending from current practice (scenario I) to older women (scenario II), the predicted incidence among the oldest age group increases from 362.0 to 421.4 (16.4%). The mortality predictions yielded by alternative scenarios are very similar: the biggest reduction is 3.0% when extending from the current practice to age group 40-65 (scenario III).

Table 4. Observed and predicted overall breast cancer incidence (1/100 000 person-years) by age at diagnosis and mortality by age at death based on alternative mass-screening scenarios. In scenario I (or A) the current practice of inviting 50-59 years old women every second year is continued, in scenario II (C) the screening service is extended from 50-59 to 50-69 and in scenario III (D) to 40-65 years old women. The scenarios are programmed to take effect in the beginning of 2005.

	Observed 2001	Scenario I 2012	Scenario II 2012	Scenario III 2012
<u>40-49</u>				
Incidence (cases)	158.1 (347)	227.9 (449)	227.9 (449)	239.0 (471)
Mortality (deaths)	10.9 (24)	9.8 (19)	9.8 (19)	8.6 (17)
<u>50-59</u>				
Incidence (cases)	313.7 (681)	423.5 (916)	423.5 (916)	406.9 (880)
Mortality (deaths)	43.1 (93)	43.6 (94)	43.6 (94)	41.9 (91)
<u>60-74</u>				
Incidence (cases)	303.4 (697)	362.0 (1064)	421.4 (1238)	397.7 (1169)
Mortality (deaths)	57.4 (131)	62.6 (184)	62.4 (183)	61.7 (181)
<u>Total</u>				
Incidence (cases)	258.3 (1725)	343.5 (2429)	368.2 (2604)	356.4 (2520)
Mortality (deaths)	37.1 (248)	42.1 (297)	41.8 (296)	40.9 (289)

6.2 Effects of separate screening program components on breast cancer incidence

The effects of separate screening program components on breast cancer incidence at different ages are quantified in Table 5 as relative risks (RR) compared to the not invited in the same age group (II). The interaction term between categorical age and screening program component had a strong impact both in localized and non-localized cancers, indicating that separate components of the screening program have different effects on the breast cancer incidence in different age groups. In general, the RR goes up during the first years of the screening rounds when the invitation and also the mammography test takes place, and declines below the baseline during the second years. This effect is larger during the first round compared to the subsequent ones and among localized cases compared to non-localized. In localized cancers, RRs during the first years of all the screening rounds tends to increase with age, whereas in non-localized cancers the trend is the opposite. Because the majority of the cancers diagnosed during the first years of the screening rounds were localized (see Table 1), the results for localized cancers dominate the results for all cancers combined.

During the second years – years between invitations – risks of invited women are clearly below the risk of non-invited women both for localized and non-localized breast cancers (Table 5). Estimates are quite stable both between localized and more advanced cases and among the age groups, except the very low estimate for age group 65-69 in non-localized cases during the 2nd years of the subsequent rounds (RR 0.19, 95%CI 0.10-0.36).

The post-screening effect up to 5 years after the program for localized cancers was elevated in the age group of 55-59, whereas it was approximately equal to the non-invited women in the older age groups (Table 5). In non-localized cases the RR estimates for the post-screening period were below 1 in the age of 60-69.

Figure 8 shows the model-based cumulative breast cancer incidence curves for (fictional) regular programs versus no screening. The term ‘regular’ indicates that women are invited every second year, and no breaks or other deviations from the official recommendations take place. In localized cancers regular invitations to screening cause 5.2% (screening up to 59) and 28.0% (screening up to 69) extra incidence when cumulated over ages 50-74 compared to situation with no mass-screening. Expanding the idealistic regular program from 50-59 to 50-69 would increase the cumulative incidence with 21.7%. In non-localized cases regular invitations up to 59 introduce a 19.8% decrease and up to 69 a 20.9% decrease in cumulative incidence compared to no screening when cumulated over ages 50-74. Up to the age of 55 the cumulative incidence of non-localized breast cancers is somewhat higher among the regularly invited ones compared to the non-invited.

Table 5: Relative risk (RR) of breast cancer and corresponding 95% confidence interval (95%CI) in each component of the screening program in women aged 50-54, 55-59, 60-64, 65-69 or 70-74 compared to non-invited women in the same age group. Cells with less than 12 000 person-years were excluded from analyses (-).

Component	Age	Localized		Non-localized		All combined *	
		RR	95%CI	RR	95%CI	RR	95%CI
1st round/1st year	50-54	2.04	(1.73, 2.41)	1.77	(1.43, 2.18)	1.92	(1.69, 2.17)
	55-59	3.03	(2.46, 3.73)	1.48	(1.13, 1.93)	2.39	(2.04, 2.80)
	60-64	3.98	(3.06, 5.17)	1.43	(0.87, 2.33)	3.07	(2.47, 3.81)
	65-69	-		-		-	
	70-74	-		-		-	
1st round/2nd year	50-54	0.64	(0.52, 0.78)	0.62	(0.49, 0.80)	0.62	(0.54, 0.73)
	55-59	0.65	(0.46, 0.90)	0.53	(0.36, 0.78)	0.62	(0.49, 0.79)
	60-64	0.59	(0.37, 0.94)	0.62	(0.37, 1.07)	0.63	(0.45, 0.89)
	65-69	-		-		-	
	70-74	-		-		-	
Subseq. rounds/1st years	50-54	1.64	(1.39, 1.94)	1.31	(1.06, 1.62)	1.50	(1.32, 1.70)
	55-59	2.28	(1.91, 2.72)	1.06	(0.86, 1.30)	1.76	(1.55, 2.00)
	60-64	2.06	(1.79, 2.36)	1.01	(0.83, 1.23)	1.61	(1.45, 1.79)
	65-69	2.63	(2.16, 3.19)	0.75	(0.52, 1.07)	1.75	(1.49, 2.06)
	70-74	-		-		-	
Subseq. rounds/2nd years	50-54	0.61	(0.50, 0.75)	0.72	(0.56, 0.92)	0.67	(0.57, 0.78)
	55-59	0.69	(0.57, 0.84)	0.52	(0.42, 0.65)	0.62	(0.54, 0.72)
	60-64	0.45	(0.37, 0.55)	0.44	(0.35, 0.56)	0.47	(0.40, 0.54)
	65-69	0.63	(0.45, 0.89)	0.19	(0.10, 0.36)	0.44	(0.34, 0.58)
	70-74	-		-		-	
Up to 5 years after	50-54	-		-		-	
	55-59	1.88	(1.28, 2.77)	0.71	(0.39, 1.29)	1.33	(0.98, 1.81)
	60-64	1.01	(0.88, 1.16)	0.81	(0.69, 0.96)	0.93	(0.84, 1.03)
	65-69	1.10	(0.96, 1.26)	0.69	(0.58, 0.81)	0.94	(0.86, 1.04)
	70-74	0.96	(0.71, 1.29)	0.99	(0.71, 1.38)	1.00	(0.82, 1.23)
More than 5 years after	50-54	-		-		-	
	55-59	-		-		-	
	60-64	-		-		-	
	65-69	0.91	(0.76, 1.08)	0.63	(0.51, 0.79)	0.80	(0.70, 0.91)
	70-74	0.80	(0.65, 0.99)	0.79	(0.63, 1.00)	0.92	(0.80, 1.06)
Break	50-54	1.76	(1.29, 2.40)	0.81	(0.47, 1.39)	1.27	(0.97, 1.66)
	55-59	1.41	(1.02, 1.94)	0.79	(0.52, 1.20)	1.13	(0.89, 1.44)
	60-64	1.32	(0.93, 1.88)	0.92	(0.57, 1.47)	1.14	(0.87, 1.49)
	65-69	-		-		-	
	70-74	-		-		-	

* Also cases with unreported stage are included

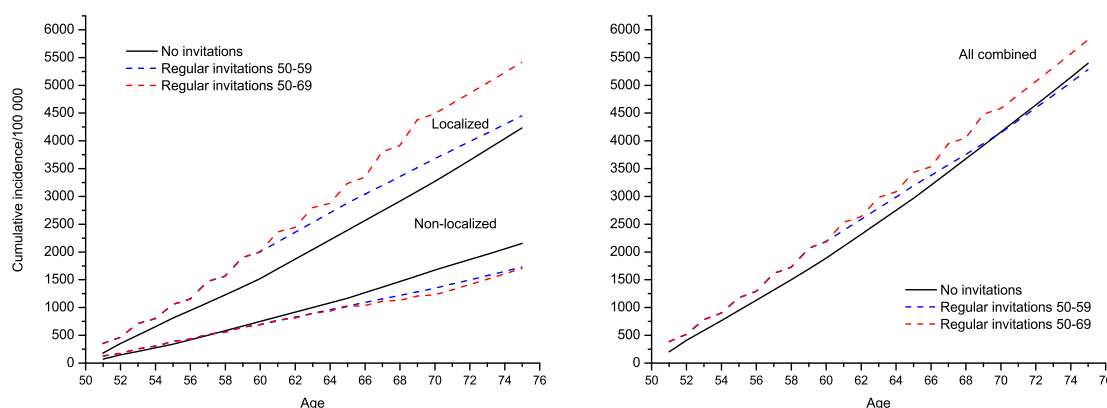


Figure 8: Model-based cumulative breast cancer incidence curves when assuming regular invitations in the age of 50-59 (blue dashed line) and regular invitations in the age of 50-69 (red dashed line) versus no screening (black solid line).

In all breast cancers combined, the cumulative incidence curve for regular invitations up to 59 reaches the expected cumulative incidence level of the non-invited at the age of 70, eleven years after the last screening (Figure 8). Extending regular screening invitations to then age of 69 increases the overall cumulative incidence up to the age of 74 with 7.9% compared to the non-invited and 10.2% compared to regular screening stopped at the age of 59.

6.3 Effects of separate screening program components on proportion of cured and life expectancy for uncured patients

Table 6 displays the estimates for proportion of cured (P) and life expectancy for uncured patients (LE) diagnosed in year 1995. The amount of uncertainty associated with the estimates is relatively high leading to wide confidence intervals and thus non-significant results. However, in cases diagnosed during the first years of the screening rounds the point estimates indicate lower (localized) or similar (non-localized) life expectancy compared to non-invited. Life expectancies for cases diagnosed during the program breaks are the lowest in all subgroups. In general, screening increased the estimated proportions of cured patients both among localized and non-localized cases and in all age groups (Table 6). The point estimates also decreased with age and were almost double in localized cases compared to women with a more advanced disease.

Table 6: Life expectancies for uncured breast cancer patients diagnosed in year 1995 at the age of 45, 55 or 65 during combined components (A, B, C, D, N) of the screening program, and corresponding proportions of cured patients. Estimates are provided with 95% confidence intervals.

	Age at diagnosis	Life expectancy (years)	Proportion of cured (%)
Localized	<u>45 years</u>		
	(A) 1 st years	7.7 (2.9, 20.2)	89.1 (77.2, 95.2)
	(B) 2 nd years and up to 5 years after	6.8 (2.3, 20.4)	87.4 (70.1, 95.3)
	(C) Break	5.1 (1.3, 20.5)	*
	(N) Not invited	9.2 (5.3, 16.0)	82.5 (74.6, 88.4)
	<u>55 years</u>		
	(A) 1 st years	12.2 (5.4, 27.6)	85.0 (72.2, 92.5)
	(B) 2 nd years and up to 5 years after	10.9 (3.8, 31.1)	82.8 (62.7, 93.2)
	(C) Break	8.2 (2.1, 31.9)	*
	(N) Not invited	14.7 (5.0, 42.7)	76.7 (55.7, 89.6)
	<u>65 years</u>		
	(A) 1 st years	11.9 (4.2, 33.7)	82.0 (61.1, 93.0)
	(B) 2 nd years and up to 5 years after	10.5 (3.5, 31.8)	79.5 (53.5, 92.9)
	(D) More than 5 years after	8.9 (2.3, 34.7)	*
	(C) Break	7.9 (1.9, 32.8)	*
	(N) Not invited	14.3 (6.3, 32.2)	72.6 (53.9, 85.7)
Non-localized	<u>45 years</u>		
	(A) 1 st years	8.3 (3.6, 19.3)	51.8 (31.0, 72.0)
	(B) 2 nd years and up to 5 years after	4.9 (2.2, 11.1)	51.0 (30.7, 70.9)
	(C) Break	7.1 (2.5, 20.3)	*
	(N) Not invited	8.2 (5.8, 11.7)	42.8 (32.3, 53.9)
	<u>55 years</u>		
	(A) 1 st years	8.9 (5.0, 15.7)	48.1 (32.9, 63.6)
	(B) 2 nd years and up to 5 years after	5.2 (2.8, 9.9)	47.3 (32.0, 63.1)
	(C) Break	7.5 (2.9, 19.4)	*
	(N) Not invited	8.7 (4.5, 16.9)	39.2 (24.5, 56.1)
	<u>65 years</u>		
	(A) 1 st years	8.3 (3.7, 18.7)	43.6 (24.2, 65.2)
	(B) 2 nd years and up to 5 years after	4.9 (2.4, 10.0)	42.8 (25.3, 62.4)
	(D) More than 5 years after	9.2 (3.0, 28.1)	*
	(C) Break	7.1 (2.6, 19.3)	*
	(N) Not invited	8.2 (5.4, 12.4)	35.0 (25.5, 45.8)

* According to the model, the same as in "2nd years and up to 5 years after"

When using a 1% unit precision to define the equality between P and $S^{(C)}$, the time needed to reach the point of cure after the diagnosis varied from 15 to 45 years in localized, and from 20 to 46 years in non-localized cases diagnosed in year 1995, depending on age and the screening component.

6.4 Uncertainty in mortality predictions related to breast cancer survival

Stage-specific confidence regions based on likelihood ratio were constructed to feed the uncertainty in the estimates from the survival model (3) into mortality predictions. Both 95% confidence regions in Figure 9 show strong negative correlation between change in LE and change in P . Therefore, even if there is substantial variation in the estimated values of LE and P (Table 6), the estimated breast cancer mortality is fairly stable as seen in Figure 10, which shows the mortality predictions 95% confidence bounds related with uncertainty in breast cancer survival estimates. In target year 2012, the confidence bounds for mortality caused by localized breast cancers are 8.9 – 10.6 cases / 100 000, and for mortality due to non-localized cancers 28.2 – 30.6 cases / 100 000.

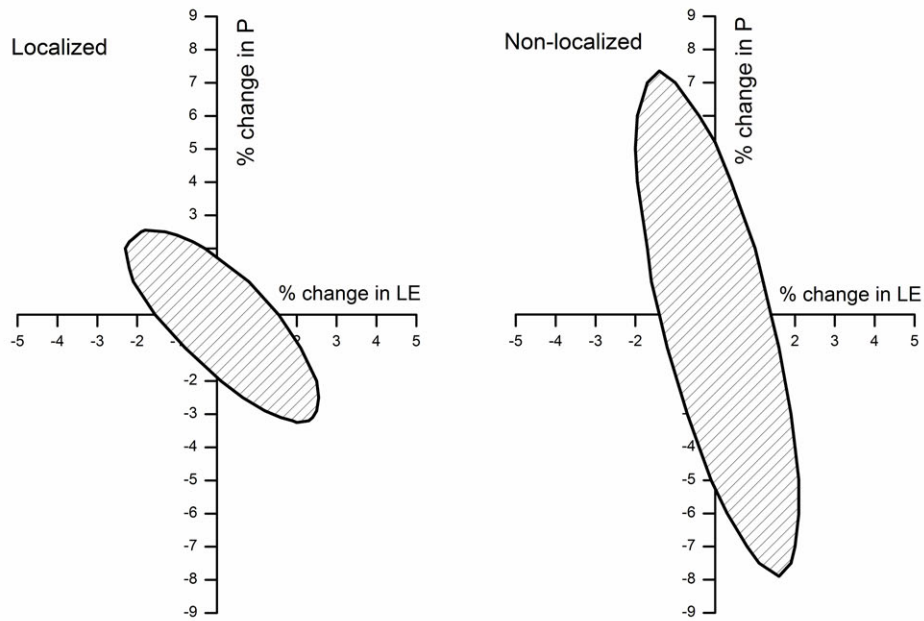


Figure 9: Likelihood-based 95% confidence regions defined by $2(L_{\max} - L) \leq \chi^2$ ($df=2$, $\alpha=0.05$), where L_{\max} is the maximum log-likelihood and L is log-likelihood obtained by altering the values of expected survival time for the uncured patients (LE) and proportion of cured patients (P).

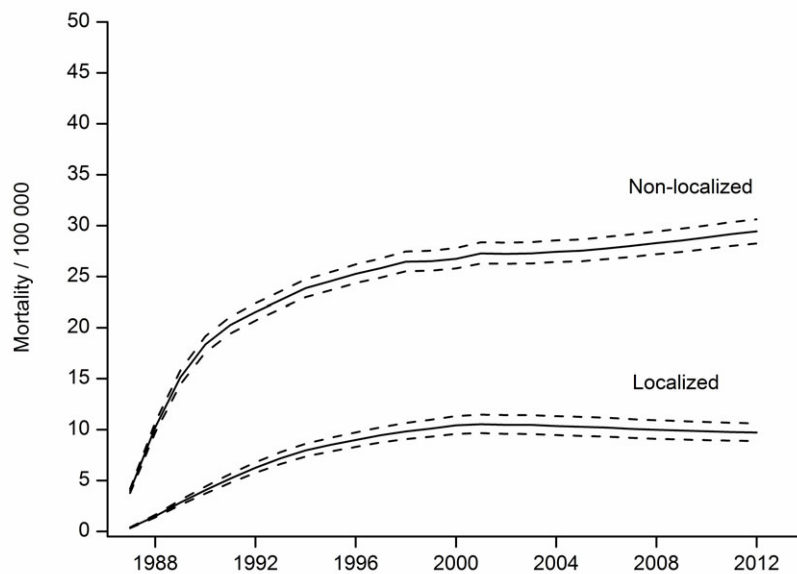


Figure 10: The breast cancer mortality predictions made under scenario I (solid line) and the 95% confidence bounds related with uncertainty in breast cancer survival estimates (dashed line).

6.5 Cancer registry vs. clinical variables as predictors for breast cancer survival

The cancer registry stage and age lost some of their importance when information about screening and clinical tumor characteristics were added in the model of breast cancer survival. The effect of stage decreased and the impact of age group changed indicating that they both acted as a surrogate for clinical variables and mammography-detection. The level of surrogacy of the cancer registry stage and age was examined by comparing the predicted numbers of breast cancer deaths yielded by different models. The predictions between the model that included only the registry stage and age (Model 1) and model including also the clinical variables (Model 3) were approximately the same, indicating that the level of surrogacy was very high. It should be noted, however, that the number of events (breast cancer deaths) was quite limited. The screen-detection, that is whether the tumor was detected by population-based mammography screening or by mammography control on own initiative, was found to be a significant factor when used as the only additive information on top of the cancer registry stage and age. The results yielded both by Cox regression and relative survival approaches were similar. Secondary analyses using the clinical data stage information instead of the stage recorded in cancer registry showed that in cases with dissimilar staging cancer registry stage gave a better picture of the breast cancer survival than the clinical stage.

7. Discussion

Detailed screening invitation data were incorporated into modeling of breast cancer incidence and survival by defining a screening variable that gives the component of the screening program by municipality, year and age. The incidence rate during the observation period 1987-2001 was then modeled using a Poisson regression approach, giving maximum likelihood estimates for the parameters, including effects of the screening program components in different age groups (I, II). These estimates, together with hypothetical scenarios of future screening policy, were then used in extrapolating the model into the future and calculating alternative incidence predictions for breast cancer up to 2012 (I). The cause-specific survival was modeled by applying a parametric mixture cure model, where the patient population was allowed to be a combination of cured and uncured patients. In addition to survival for the uncured, effects of covariates were also incorporated through the proportion of cured patients and shape and scale parameters of the survival time gamma distribution. The patients' risk to die from other causes than breast cancer was allowed to differ from that of a corresponding general population group and to depend on age and follow-up time. The incidence, cause-specific and other cause survival were then combined and alternative short-time breast cancer mortality predictions based on hypothetical scenarios of future screening policy were calculated (III). Finally, the power of cancer registry variables stage and age as predictors for breast cancer survival was established by showing that they act as surrogates for clinical tumor characteristics (IV).

Some strong assumptions concerning the screening data were made: First, only invitations to the screening test were known, not the real attendance activity. If the calendar year of woman's invitation and of her breast cancer diagnosis were the same, it was assumed that the cancer was diagnosed as a result of mass-screening. This assumption is strong but seems to hold surprisingly well in our study. For extrapolations it means that an invitation pattern is regarded to result in a certain level of attendance which is assumed to be stable in the future. On the other hand, the attendance level is known to be very high among Finnish women, and there is no such a reason within sight that would change this behavior. Secondly, year of diagnosis together with birth year was used to link the cancer cases with the invitation data. Mammographic screening is, however, a multiple-step process, starting with the initial mammogram and leading, if positive, to a series of more detailed investigations. Because the inevitable time lag between invitation to screening test and the diagnosis, particularly women invited in the end of the calendar year might have their diagnosis not until the next year. There is no doubt this introduces some bias in the estimates, and one could think that the true differences in the screening effect sizes between the first actual screening year and the second year of the round are larger since some of the diagnoses resulting from screening have 'slipped' over to the next calendar year.

A potential confounder in estimating the effects of screening both on breast cancer incidence and survival is the opportunistic mammography screening performed outside the official mass-screening program. At least for the time being, there are no register data available on these activities in Finland. Using a new approach based on proportion of DCIS, Weedon-Fekjær has suggested that the level of opportunistic screening among 50-69 years old women included in the Norwegian breast cancer screening program is 17.9% with a 95% CI of 14.4 – 23.7 [41]. In Finland, opportunistic screening includes also invitational mammographic tests paid by women themselves, which has become a commonplace practice among municipalities; in year 1997 approximately 24% of women aged 60-69 were offered this self-paid screening service in Finland [42]. At that time, there were no invitations to self-paid tests for women younger than 60. However, opportunistic screening confounds the estimates of screening efficacy – not the assessment of policy changes, assuming that the level and quality of opportunistic screening remains similar to the situation during the observation period.

The university hospital region was included as a factor in the incidence models (I, II), but left out from the survival model because it did not improve the fit (III). This indicates that even if there is regional variation in baseline breast cancer incidence in Finland, survival time from breast cancer is not a function of an oncological region. This is in line with the fact that almost all cancer patients in Finland are treated in public hospitals which are obliged to follow common agreed guidelines, leading to quite homogenous treatment practices of cancer patients within the whole country. Furthermore, the division by university hospital region is too general for studying geographical differences in the performance of the Finnish screening program because these differences often originate from dissimilarities in screening practices between municipalities. To take into account the variation in incidence between municipalities, we performed an additional analysis where the municipality was included in the model and defined as a random variable while the other predictors were held fixed [43]. As expected, this resulted in some increase of the standard errors, but changes from the original results were minor.

The effects of mass-screening were clearly visible in the estimates of the incidence of both localized and non-localized cancers during the active part of the programme (Table 5). The relative risk went up during the first years of the screening rounds when the invitation and screening test (in most of the cases) took place, and declined below the baseline during the second years, implying that screening increases the immediate incidence of breast cancers. This effect was larger during the first round compared to the later ones, mainly because of the detection of the so called ‘prevalence pool’, which includes lesions with a long sojourn time [10]. In localized cancers, the relative risk during the first years of the screening rounds tended to increase with age, whereas in non-localized cancers the trend was the opposite. This may be linked with the fact that validity of screening is higher in older than in younger age groups because of the changes in breast tissue with age: a breast cancer is

more readily detected in a fatty breast than in young women with dense breasts [10, 44]. For this reason the test detects localized cancers before they spread and become non-localized better in the older than in the younger age groups.

The point of statistical cure, defined as the point in follow-up time when the breast cancer survival curve reaches the level of P , describes the amount of time needed until the only patients that are left alive (from breast cancer) among the study population are the cured ones, in other words, the time needed until all patients in the 'bound to die' population have died. For example, among patients diagnosed with a non-localized cancer in year 1995 at the age of 55 during the first years of the screening round, it would take 36 years until the excess mortality associated with breast cancer has disappeared (data not shown). In a patient's real life, the recovery from the disease takes place some time after the treatment. Therefore, the results for statistical cure point have no direct interpretation in real life and are presented more in a demonstrative sense than to be taken as hard facts. Also precision used for calculating the point of cure turned out to be very crucial due to long and flat tails of the survival distributions.

Estimating the proportion of cured is a matter of identifiability between the cure fraction and the mean survival time of the uncured. Distinguishing between delayed mortality and cure may be a problem especially in patients with a long survival time like in breast cancer. However, from extrapolation point of view this makes no difference and therefore identifiability problems do not invalidate the use of cure models when making mortality predictions into the near future. Besides, even if the model fits the observed follow-up period well, estimates for P may be inappropriate as they are based on extrapolation of the parametric distribution beyond the range of the data [31].

According to our results, screening increases the proportion of cured patients compared to the non-invited among localized cases but decreases the corresponding life expectancy for the uncured patients (Table 6); screening removes some patients with relatively good prognosis from the group of fatal cases to the group of cured ones. Since this is not happening in the group of non-invited women, their prognoses are better on average.

The small changes in breast cancer mortality predictions after extending the screening program may be surprising (Figure 7). In fact, all the predictions fall within the confidence limits presented in Figure 10, except the prediction based on scenario III in non-localized cases being at the lower borderline. However, one limitation in our analysis is that the results are restricted to ages below 75. As both the lead and survival times in breast cancer are long [45], it might be that benefits of extending the program to 69 or to 65 years old women would be visible only above the age of 74. The results can also partly be explained by invitational mammographic tests paid by women themselves, which are known to be common among Finnish women above 60. Even so, it is unlikely that these limitations

would explain the similarities between alternative predictions in full. It could be that the current program inviting women between ages 50-59 complemented with the (unavoidable) opportunistic screening is sufficient, and potential decreases in stage-specific mortality obtained by extending the program to older women are also achieved by effective treatments [24]. As medical treatments progress, it may be that the prognoses for cancers detected in elderly women only when they give clinical symptoms are as good as for those preclinical cases detected by screening. Moreover, since most of the breast cancer deaths are due to metastases to other parts of the body, a critical question is when the disease metastasizes, and whether mammography screening is able to bring the diagnosis and treatment earlier in time enough to avoid the spreading of the disease.

To provide some comparisons, according to predictions based on Finnish Cancer Registry data from year 1955-2003 there would be a total of 3488 new breast cancer cases in Finnish female population in age group 40-74 in year 2012 (Finnish Cancer registry); knowing that the study population covers 59.0% of 40-74 years-old women in Finland, the adjusted prediction for year 2012 would be 2058 cases. Further, to make figures comparable in situ tumors have to be added; if their proportion is assumed to be 3.7% as in the study database (see section 4.5), we end up with 2134 predicted new cases. The predictions for the age group 40-74 for year 2012 presented in this study (I) vary from 2234 to 2604 breast cancer cases between different scenarios, so they are a bit higher than the ‘official’ Finnish predictions.

Since breast cancer mortality was calculated as a combination of different, independently fitted models, a common variance for the parameters in mortality model could not be directly calculated. As a solution, an approach where only the uncertainty related to breast cancer survival was taken into account was presented, leaving out the contributions of incidence and other cause mortality models. It was shown that despite the wide confidence intervals of the survival model components life expectancy and P , the uncertainty related to breast cancer survival was reasonable leading to fairly stable mortality estimates. This is due to the strong negative correlation between the changes in those components. As in Hakulinen and Dyba [46], the uncertainty related to the incidence analyzed with Poisson regression could be approximated using the delta method based on Taylor series expansion. Also bootstrapping or Bayesian approaches could be useful. The variation in the other cause mortality estimates consists of the error related to the population death probabilities and the error due to the estimates of correction coefficient r ; the former can be assumed to be small and the latter was found out to be reasonable (III). In conclusion, the true prediction intervals of breast cancer mortality predictions are wider than the presented ones, the increase in uncertainty originating mainly from the incidence model. Also the precision of predictions based on individual data is clearly an issue subject to further study.

During the model-building process of the cure model, the reduction of the number of parameters became unavoidable in order to provide stable estimates for each subgroup. As a consequence, components of the screening program had to be combined. This is a good example of practical limitations met by the researchers aiming to optimal models. Even if the Finnish cancer registry is one of the biggest in the world, there were not enough cases (or deaths) to provide the planned survival analysis because of the relatively small population and lack of screening information from 40% of the municipalities. The lack of power may prevent the detection of the modest differences. This emphasises the great value of nation wide cancer registries with high quality data and good coverage in countries with large populations for cancer epidemiological research.

Economical and administrative reasons have led to the current situation where the Finnish municipalities are increasingly contracting the mass-screening to private institutions instead of the Cancer Society screening centres. Despite mandatory sending of the screening information to the Mass Screening Registry, arrangements are always a subject to negotiations and variation in the field of operators will cause considerable problems in maintaining the central register of screening information. This raises concerns about the future of population based studies evaluating the Finnish national mass-screening program.

8. Conclusions

- The impacts of screening policy changes are clearly visible on the predictions for breast cancer incidence in age group 40-74, while the effects on short time mortality predictions are minor. Furthermore, the changes would influence more localized than non-localized incidence rates.
- Given the quality and attendance rate of the current screening process, effective cancer treatments, and the current rate of confounding caused by opportunistic screening, this study gives little support that extending the Finnish mass-screening service to women older than 60 would decrease the breast cancer mortality up to age 74 in the near future. The results show no major differences between mortality predictions up to 2012 yielded by alternative future scenarios of the screening policy: Any policy change would have at the most a 3.0% reduction on overall breast cancer mortality compared to continuing the current practice in the near future. The reduction on mortality resulting from cases diagnosed as non-localized would be at the most 4.5%. Extensions of the program would have no significant impact on mortality caused by cases diagnosed as localized.
- Extending the screening service would increase the localized but decrease the non-localized breast cancer incidence. The biggest increase in the incidence of localized cases (13.5%) is introduced by extending the service from age group 50-59 to 50-69, and the biggest decrease in non-localized cases (9.7%) would be achieved by extending to the age group 40-65.
- Clinical tumor characteristics are not necessarily needed when making breast cancer survival predictions based on a population-based cancer registry: The cancer registry stage and age act as surrogates for clinical information.
- Providing estimates about the uncertainty related to mortality predictions based on model combinations and individual data is an issue that needs further study.

References

1. Ferlay J, Autier P, Boniol M, Heanue M, Colombet M, Boyle P. Estimates of the cancer incidence and mortality in Europe in 2006. *Annals of Oncology* 2007; 18: 581-592.
2. Stewart BW, Kleihues P. World cancer report. Lyon: IARC Press, 2003.
3. Hakama M, Pukkala E, Söderman B, Day N. Implementation of screening as a public health policy: issues in design and evaluation. *Journal of Medical Screening* 1999; 6: 209-216.
4. Anttila A, Koskela J, Hakama M. Programme sensitivity and effectiveness of mammography service screening in Helsinki, Finland. *Journal of Medical Screening* 2002; 9: 153-8.
5. Parvinen I, Helenius H, Pylkkänen L, Anttila A, Immonen-Räihä P, Kauhava L, Räsänen O, Klemi PJ. Service screening mammography reduces breast cancer mortality among elderly women in Turku. *Journal of Medical Screening* 2006; 13: 34-40.
6. Sarkeala T, Heinävaara S, Anttila A. Organised mammography screening reduces breast cancer mortality: A cohort study from Finland. *International Journal of Cancer* 2008; 122: 614-619.
7. Anttila A, Sarkeala T, Hakulinen T, Heinävaara S. Impacts of the Finnish service screening programme on breast cancer rates. *BMC Public Health* 2008; 8: 38.
8. Hankinson S, Husebye E. Breast Cancer. In: Adami H, Hunter D, Trichopoulos D, editors. Textbook of Cancer Epidemiology. Oxford University Press, 2002; 301-339.
9. Last J. A Dictionary of Epidemiology. New York, Oxford University Press, p.52.
10. IARC Working Group on the Evaluation of Cancer-Preventive Strategies. Breast cancer screening. IARC Handbooks of Cancer Prevention, vol. 7. Lyon: IARC Press, 2002.
11. Heinävaara S, Hakulinen T. Predicting the lung cancer burden: Accounting for selection of the patients with respect to general population mortality. *Statistics in Medicine* 2006; 25: 2967-80.
12. De Angelis R, Capocaccia R, Hakulinen T, Söderman B, Verdecchia A. Mixture models for cancer survival analysis: application to population-based data with covariates. *Statistics in Medicine* 1999; 18: 441-54.

13. Phillips N, Coldman A, McBride ML. Estimating cancer prevalence using mixture models for cancer survival. *Statistics in Medicine* 2002; 21: 1257-70.
14. Nyström L, Andersson I, Bjurstam N, Frisell J, Nordenskjöld B. Long term effects of mammography screening: updated overview of the Swedish randomized trials. *Lancet* 2002; 359: 909-919.
15. Sarkeala T, Heinävaara S, Anttila A. Breast cancer mortality with varying invitational policies in organised mammography. *British Journal of Cancer* 2008; 98: 641-645.
16. Moss S, Cuckle H, Evans A, Johns L, Waller M, Bobrow L. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomized controlled trial. *Lancet* 368:2053-60.
17. De Koning H, van Ineveld B, van Oortmarssen G, de Haes J, Collette H, Hendriks J, van der Maas P. Breast cancer screening and cost-effectiveness; Policy alternatives, quality of life considerations and the possible impact of uncertain factors. *International Journal of Cancer* 1991; 49:531-537.
18. Hakulinen T, Pukkala E. Future incidence of lung cancer: forecasts based on hypothetical changes in the smoking habits of males. *International Journal of Epidemiology* 1981; 10: 233-240.
19. Møller B, Fekjær H, Hakulinen T. Prediction of cancer incidence in the Nordic countries: empirical comparison of different approaches. *Statistics in Medicine* 2003; 22: 2751-2766.
20. Dyba T, Hakulinen T. Do cancer predictions work? *European Journal of Cancer* 2008; 44: 448-453.
21. Pukkala E, Dyba T, Hakulinen T, Sankila R. Syövän ilmaantuvuus, syöpäpotilaiden ennuste ja syöpäkuolleisuus. In: Hakama M, Hakulinen T, Johansson R, Rautalahti M, Vertio H. Syöpä 2015. Suomen Syöpäyhdistyksen julkaisuja nro 71. Suomen Syöpäyhdistys, Helsinki 2006.
22. Møller B, Weedon-Fekjær H, Hakulinen T, Tryggvadóttir L, Storm HH, Talbäck M, Haldorsen T. The influence of mammographic screening on national trends in breast cancer incidence. *European Journal of Cancer Prevention* 2005; 14(2): 117-128.
23. Møller B, Fekjær H, Hakulinen T, Tryggvadóttir L, Storm HH, Talbäck M, Haldorsen T. Prediction of cancer incidence in the Nordic countries up to the year 2020. *European Journal of Cancer Prevention Suppl* 2002; 11: 1-96.

24. Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, Mandelblatt JS, Yakovlev AY, Habbema JDF, Feuer EJ. Effect of screening and adjuvant therapy on mortality from breast cancer. *The New England Journal of Medicine* 2005; 353: 1784-92.
25. Sasieni P, Adams J. Analysis of cervical cancer mortality and incidence data from England and Wales: evidence of a beneficial effect of screening. *Journal of the Royal Statistical Society A* 2000; 163: 191-209.
26. Holford, T. The estimation of age, period and cohort effects for vital rates. *Biometrics* 1983; 39: 311-324.
27. Clayton D, Schifflers E. Models for temporal variation in cancer rates. I: Age-period and age-cohort models. *Statistics in Medicine* 1987; 6: 449-467.
28. Clayton D, Schifflers E. Models for temporal variation in cancer rates. II: Age-period-cohort models. *Statistics in Medicine* 1987; 6: 469-481.
29. Boag J. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society B* 1949; 11: 15-44.
30. Yu B, Tiwari RC, Cronin KA, Feuer EJ. Cure fraction estimation from the mixture cure models for grouped survival data. *Statistics in Medicine* 2004; 23: 1733-1747.
31. Lambert PC, Thompson JR, Weston CL, Dickman PW. Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics* 2007; 8(3):576-594.
32. Hakulinen T, Kenward M, Luostarinen T, Oksanen H, Pukkala E, Söderman B, Teppo L. Cancer in Finland in 1954-2008: Incidence, mortality and prevalence by region. Finnish Foundation for Cancer Research, Helsinki 1989.
33. Verdecchia A, De Angelis G, Capocaccia R. Estimation and projections of cancer prevalence from cancer registry data. *Statistics in Medicine* 2002; 21: 3511-3526.
34. Sarkeala T, Anttila A, Forsman H, Luostarinen T, Saarenmaa I, Hakama M. Process indicators from ten centres in the Finnish breast cancer screening programme from 1991 to 2000. *European Journal of Cancer* 2004; 40: 2116-2125.
35. Berrino F, Capocaccia R, Coleman MP *et al.* Survival of cancer patients in Europe: the EURO CARE-3 study. *Annals of Oncology* 2003; 14 (Supplement 5): 9-13.
36. Parvinen I, Helenius H, Pylkkänen L, Anttila A, Immonen-Räihä P, Kauhava L, Räsänen O, Klemi PJ. Service screening mammography reduces breast cancer mortality among elderly women in Turku. *Journal of Medical Screening* 2006; 13: 34-40.

37. Finnish Cancer Registry. *Cancer in Finland 2004 and 2005*. Cancer Society of Finland Publication No. 72, Helsinki, 2007; 13.
38. SAS Institute Inc. SAS/STAT® User's Guide, Version 8. SAS Institute Inc.: Cary, NC, 1999.
39. Coleman MP, Esteve J, Damiecki P, Arslan A, Renard H Trends in cancer incidence and mortality. IARC Scientific Publications No 121. Lyon: IARC, 1993.
40. Dickman PW, Sloggett A, Hills M, Hakulinen T. Regression models for relative survival. *Statistics in Medicine* 2004; 23: 51-64.
41. Weedon-Fekjær H. Modelling breast cancer incidence, progression and screening test sensitivity using screening data. Faculty of Medicine, University of Oslo No.570, 2007; 57-59.
42. Saarenmaa I, Salminen T, Varonen H, Fredriksson M, Sintonen H, Mäkelä M. FinOHTAn raportti 16: Rintasyöpäseulonnan laajentamisen vaikutukset. Helsinki: STAKES, 2000; 23.
43. McCulloch CE, Searle SR. Generalized, Linear, and Mixed Models. Wiley, New York, 2000.
44. Sarkeala T, Hakama M, Saarenmaa I, Hakulinen T, Forsman H, Anttila A. Episode sensitivity in association with process indicators in the Finnish breast cancer screening programme. *International Journal of Cancer* 2006; 118:174-179.
45. Paci E, Miccinesi G, Puliti D, Baldazzi P, De Lisi V, Falcini F, Cirilli C, Ferretti S, Mangone L, Finarelli AC, Rosso S, Segnan N, Stracci F, Traina A, Tumino R, Zorzi M. Estimate of overdiagnosis of breast cancer due to mammography after adjustment for lead time. A service screening in Italy. *Breast Cancer Research* 2006; 8: R68.
46. Hakulinen T, Dyba T. Precision of incidence predictions based on Poisson distributed observations. *Statistics in Medicine* 1994; 13:1513-1523.